

In Press, JEP: General, 17/February/2019

Cognitive Control and Capacity for Prospective Memory in Complex Dynamic Environments

Boag¹, R., Strickland², L., Heathcote², A., Neal³, A., & Loft¹, S.

¹The University of Western Australia

²The University of Tasmania

³The University of Queensland

Word count: 18,322

Author Note

This research was supported by Discovery Grant DP160101891 from the Australian Research Council awarded to Heathcote, Neal, & Loft. Part of this research was presented at the Australasian Mathematical Psychology Conference (2017). Correspondence concerning this article should be addressed to Russell Boag, School of Psychological Science, University of Western Australia, Crawley, WA 6009, Australia (email: russell.boag@research.uwa.edu.au).

Abstract

Performing deferred actions in the future relies upon Prospective Memory (PM). Often, PM demands arise in complex dynamic tasks. Not only can PM be challenging in such environments, the processes required for PM may affect the performance of other tasks. To adapt to PM demands in such environments, humans may use a range of strategies, including flexible allocation of cognitive resources and cognitive control mechanisms. We sought to understand such mechanisms by using the Prospective Memory Decision Control (Strickland et al., 2018) model to provide a comprehensive, quantitative account of dual task performance in a complex dynamic environment (a simulated air traffic control conflict detection task). We found that PM demands encouraged proactive control over ongoing task decisions, but that this control was reduced at high time pressure to facilitate fast responding. We found reactive inhibitory control over ongoing task processes when PM targets were encountered, and that time pressure and PM demand both affect the attentional system, increasing the amount of cognitive resources available. However, as demands exceeded the capacity limit of the cognitive system, resources were reallocated (shared) between the ongoing and PM tasks. As the ongoing task used more resources to compensate for additional time pressure demands, it drained resources that would have otherwise been available for PM task processing. This study provides the first detailed quantitative understanding of how attentional resources and cognitive control mechanisms support PM and ongoing task performance in complex dynamic environments.

Keywords: prospective memory; cognitive control; capacity; resource theory; linear ballistic accumulator model

Event-Based Prospective Memory (PM) refers to the ability to remember to perform deferred task actions when a particular stimulus or event is encountered in the future (Einstein & McDaniel, 1990). Both at work, and in everyday life, people have to manage PM demands in complex dynamic environments. For example, air traffic controllers, pilots, surgeons, nurses and emergency response personnel often have to defer taking a critical action in order to deal with higher priority tasks (Dismukes, 2012; Grundgeiger, Sanderson, MacDougall, & Venkatesh, 2010; Loft, 2014). This means that these experts need to be able to manage the demands of the PM task alongside the demands of ongoing tasks, which can be difficult when those ongoing tasks are highly demanding, and there is significant time pressure. Errors of PM (i.e., failing to execute a deferred intention) can have serious consequences both at work and in everyday life, making it important to understand how people coordinate the demands of PM with that of ongoing tasks.

Over the past 30 years, a great deal has been learnt about PM using tightly controlled laboratory paradigms. One of the most consistent findings has been that PM interferes with ongoing task performance. In a typical PM study, participants complete a relatively simple ongoing decision-making task (e.g., lexical decision, categorization) by itself, or with a concurrent PM task. For example, the ongoing task might be a lexical decision task, and the PM instruction might be to press an alternative key if presented a category of word. Under PM load, individuals are slower to respond to items which do not require a PM response, a phenomenon known as *PM cost* (e.g., Einstein & McDaniel, 2005; Hicks, Marsh, & Cook, 2005; Loft & Yeo, 2007; Smith, 2003). It is typically assumed that PM costs result from *resource bottlenecks*, in which the two tasks compete for the same limited-capacity cognitive resources (Navon & Gopher, 1979; Norman & Bobrow, 1976), and that the extent of PM cost reflects the extent to

which resources are shared between the PM and ongoing task (Einstein & McDaniel, 2005; Scullin, McDaniel, & Shelton, 2013; Smith, 2003).

It has long been recognized that dual tasks may also engage *cognitive control* processes, which adapt the cognitive system to meet specific task demands (Braver, 2012; Braver, Barch, Gray, Molfese, & Snyder, 2001; Miller & Cohen, 2001; Miyake et al., 2000). Recently Strickland, Loft, Remington and Heathcote (2018) proposed a quantitative model of PM, Prospective Memory Decision Control (PMDC), which is able to disentangle the contributions of capacity sharing and cognitive control to PM and ongoing task performance. Applying PMDC to basic laboratory tasks indicated that the PM cost observed in the standard laboratory paradigm arises because participants set higher decision thresholds for the ongoing task when under PM load. This suggests that, rather than cost resulting from capacity sharing, it results from participants delaying ongoing task decisions to avoid pre-empting the PM decision (also see for example Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland, Heathcote, Remington, & Loft, 2017; Strickland et al., 2018). Strickland et al.'s model revealed that cognitive control, and not capacity sharing with ongoing tasks, was critical to PM accuracy.

We do not yet know whether the findings that PM relies upon cognitive control over ongoing task decisions, and does not affect capacity for the ongoing task, reflect essential characteristics of PM, or are more specifically tied to the types of tasks often used in the study of PM in the laboratory. The ongoing tasks used in laboratory studies of PM are often relatively simple and may not fully engage cognitive capacity. In this case, idle resources could be recruited when PM demands are added without reducing ongoing task capacity. In everyday life, however, PM can occur in the context of primary tasks that are highly demanding, and which generate significant time pressure. In these types of environments, the ongoing task may entirely

occupy capacity such that additional capacity required for the PM task must be drawn from ongoing task resources. Furthermore, simple laboratory paradigms may omit aspects of PM that are important in more representative conditions. That is, the relative contribution of cognitive control versus attentional resources in supporting PM in more complex dynamic environments tasks may differ from those previously identified for basic PM tasks.

The aim of the current work is to investigate the potential of PMDC to explain how individuals deal with PM demands and time constraints in complex dynamic environments. To do this, we have participants perform an air traffic control conflict detection task while having to remember to carry out a deferred action. The conflict detection task was used because it is a prototypical example of a broad range of work tasks that require people to remember to perform deferred task actions while making judgements about objects moving on task displays (e.g., maritime surveillance, train control, unmanned vehicle control, air battle management, submarine track management; Dismukes, 2010, 2012; Loft, 2014). We fit PMDC to the accuracy and response time (RT) data from both the ongoing task and PM task simultaneously, enabling inferences to be drawn regarding the cognitive processes that drive both tasks, which prior to Strickland et al. (2018) had not previously been achieved in PM research. Our model provides the first detailed quantitative understanding of the attentional resources and cognitive control mechanisms that support PM and ongoing task performance in complex dynamic task environments.

Prospective Memory Decision Control

PMDC belongs to the broad class of *evidence accumulation models* (e.g., Brown & Heathcote, 2008; Ratcliff, 1978), which assume that decisions are made by accumulating evidence from the environment until a threshold amount is reached. Evidence accumulation

models provide a full quantitative account of both RT distributions and the accuracy of decisions in benchmark empirical phenomena occurring in simple decision tasks, including trade-offs between speed, accuracy and response bias, and subtle interactions between the speed of correct and error responses (see Rae, Heathcote, Donkin, Averell, & Brown, 2014, for perceptual-, lexical-, and memory-based examples).

PMDC is an instance of the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008), which formalizes decision making as a process of evidence accumulation among independent racers. The PMDC model uses three racing LBA accumulators; two correspond to the ongoing task responses and a third corresponds to the PM task response. Correct PM responses (PM hits) occur on PM trials when the PM accumulator reaches threshold before either of the ongoing task accumulators. PM misses occur when one of the ongoing task accumulators finishes before the PM accumulator. In Figure 1, we depict how this would apply in a relative trajectory judgment (conflict detection) task (e.g., deciding whether two aircraft are in conflict) with an additional PM requirement. There are three possible response alternatives, which correspond to indicating that the stimulus is a *conflict*, a *non-conflict*, or a *PM target*. Each response choice has its own accumulator that accrues evidence linearly (arrows in Figure 1), starting from points (uniformly distributed over the interval 0- A) representing random trial-to-trial biases. Evidence increases over time in a linear manner until the total in one accumulator reaches its response threshold (b). To avoid circularity, b is assumed to be set prior to the trial – if it could be altered contingent on the identity of the stimulus there would be no need to make a decision. The rate of accumulation – which varies normally from trial to trial with mean v and standard deviation sv – corresponds to the strength of evidence for a choice, and it is determined both by the quality of information available to the decision maker and the degree to which he or

she attends to that information. The time for non-decision processes (e.g., stimulus encoding and response selection, ter) is estimated by the difference between the observed RT and decision time (i.e., the time that evidence in the winning accumulator first equals the threshold).

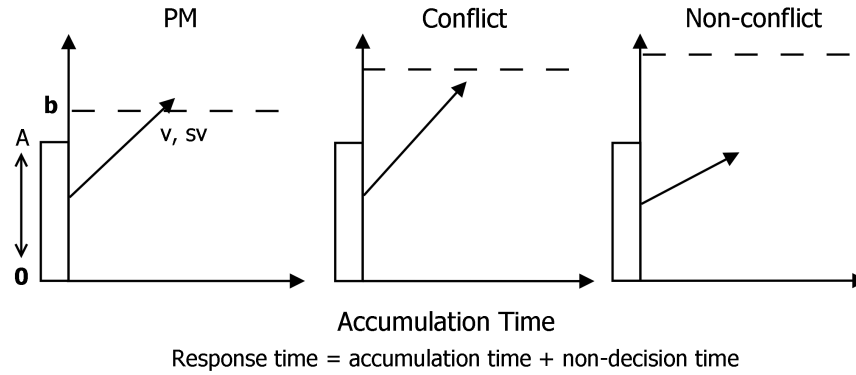


Figure 1. An LBA model of a PM task with a concurrent conflict detection task. Evidence for each response is initially drawn from a uniform distribution on the interval $[0, A]$. Over time, evidence accumulates towards each response at rates drawn from independent normal distributions with mean v , and standard deviation sv . The first accumulator to reach its threshold, b , determines the overt response. We refer in our results to B , which is $b - A$, where $B > 0$ and so $b > A$. Total RT is determined by accumulation time plus non-decision time.

The aim of fitting the PMDC model is to measure the psychological quantities that underlie performance, and to ascertain what those quantities suggest about resource allocation and cognitive control. PMDC instantiates the *proactive* and *reactive* cognitive control mechanisms specified by Braver's (2012) dual-mechanisms theory and measures the overall availability of resources, as well as the degree to which resources are shared between ongoing task processing and PM task processing. We now review how PMDC instantiates these mechanisms, and what has been found in the basic laboratory work with PMDC and other evidence accumulation models.

Proactive control. Proactive control refers to processes used to "bias attention, perception and action systems in a goal-driven manner" (Braver, 2012, p. 2). Consistent with

current theoretical (Bugg, McDaniel, & Einstein, 2013) and empirical PM research (e.g., Ball & Brewer, 2018), proactive processes are deployed deliberately, in advance of the target stimulus so that they are already active when the target stimulus is encountered. Proactive control should thus be observable in differences between the latent processes underlying responses to (non-PM) ongoing task items in PM blocks and control blocks. Under PMDC, participants proactively control task demands by raising their thresholds in PM blocks, so that on PM trials the ongoing task accumulators are less likely to complete before the PM accumulator, thereby reducing the probability of a PM miss. This claim, originally made by the delay theory of PM cost (Heathcote et al., 2015), follows from the fact that in evidence accumulation models, thresholds are the locus of *a priori* strategies that drive mechanisms such as the speed-accuracy trade-off (Liu & Watanabe, 2012) and response biases (Donkin, Brown & Heathcote, 2011; Mulder, Wagenmakers, Ratcliff, Boekel & Forstmann, 2012).

Every evidence accumulation modeling study to date that has compared ongoing task performance between control and PM blocks has found elevated thresholds in the latter (Anderson, Rummel & McDaniel, 2018; Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Horn et al., 2011, 2013; Strickland et al., 2017, 2018), consistent with proactive control. Further implicating control, Strickland et al. (2018) showed that participants increased their ongoing task thresholds further when instructed that the PM task was important. Moreover, participants also exerted control over PM thresholds (i.e., the evidence required to make a PM response) as a function of PM demands. PM thresholds were lower when the PM task was important versus not important, and lower when the features that need to be assessed to detect PM targets were focal to ongoing task processing requirements, as compared to non-focal,

indicating that ongoing task thresholds and PM thresholds are both important loci of proactive cognitive control strategies.

Reactive control. In contrast to proactive control, reactive control refers to automatic, stimulus-driven cognitive mechanisms that are deployed to influence responding "only as needed, in a just-in-time manner" (Braver, 2012, p. 2). Reactive control mechanisms are 'bottom-up' processes and are assumed to operate automatically in response to inputs signalling the critical event. As such, reactive control processes relevant to PM occur specifically on PM trials. PMDC's reactive control mechanisms are depicted in Figure 2.

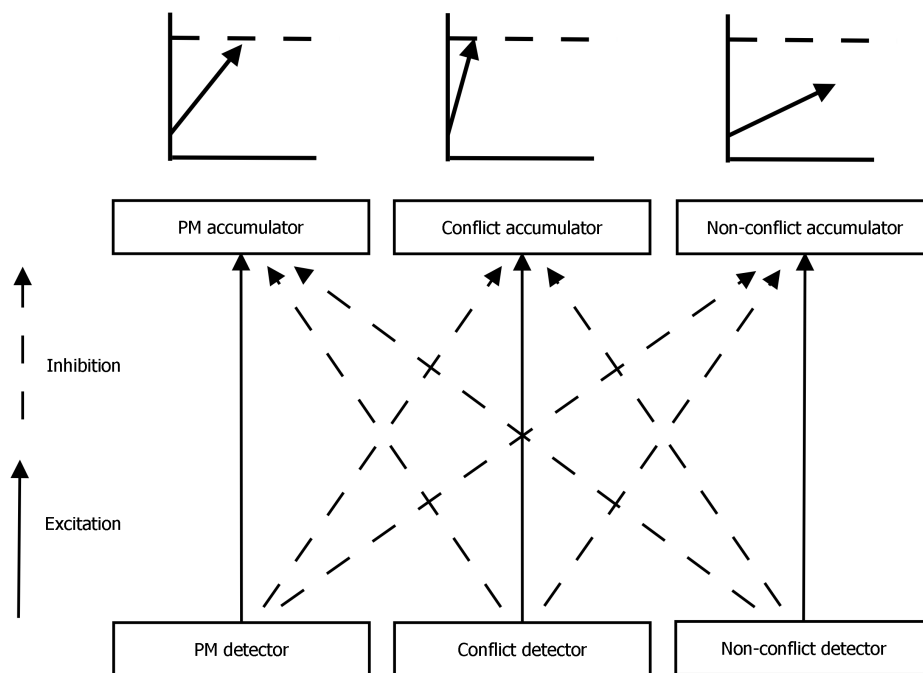


Figure 2. Reactive control of accumulation rates in PMDC, using the example of an air traffic control conflict detection task in which participants have a concurrent PM task. The encoding process includes detectors (rectangles) for each possible response to the task: 'conflict', 'non-conflict', and 'PM'. The detectors receive input from stimulus features. Output from the detectors can directly increase (solid lines) the input to the corresponding evidence accumulator (excitation) or reduce (dashed lines) the input to competing accumulators (inhibition).

PMDC claims that as PM stimulus inputs are processed on PM trials, stimulus features consistent with PM excite the PM accumulator, increasing accumulation speed, and inhibit other accumulators (reactive inhibition), decreasing their accumulation speed. Thus, in addition to PM accumulation being faster on PM trials than non-PM trials, accumulation towards ongoing task decisions should be slower on PM trials than non-PM trials.¹ The latter is the hallmark of reactive control. In line with this, Strickland et al. (2018) found that ongoing task accumulation rates were inhibited (reduced) on PM trials, and that this response competition brought about by reactive control was critical to PM performance in basic PM laboratory paradigms. This converges with other theoretical work (Bugg et al., 2013) and neurological data (McDaniel, LaMontagne, Beck, Scullin, & Braver, 2013) implicating reactive control in PM, as well as broader approaches to human error that contend that response inhibition is required for atypical task responses to be able to compete for retrieval with task responses more strongly associated with common environmental cues (Norman, 1981; Reason, 1990).

Capacity Sharing. PMDC measures capacity with the accumulation rate parameters, as have other models for measuring ongoing task capacity (Boywitt & Rummel, 2012; Horn et al., 2011). Rates correspond to capacity because they estimate processing speed, which the majority of attention theory assumes should vary in proportion to the attentional capacity available (e.g.,

¹ Readers familiar with the *intention superiority effect*, in which ongoing task RTs are faster on PM target trials than non-PM trials, may wonder whether reactive *inhibition* of rates on PM target trials is incompatible with the *faster* RTs observed in the intention superiority literature (e.g., Marsh, Hicks, & Watson, 2002). However, on PM target trials, accumulators for the ongoing task responses must compete with a much faster PM response accumulator. Overt ongoing task responses on PM trials are therefore more likely to be fast errors that outpace the PM accumulation process, a phenomenon known as statistical facilitation (Raab, 1962). As such, fast PM miss RTs are not incompatible with lower PM miss accumulation rates. In fact, we find both effects (see Table S5 in the supplementary materials and Model Analysis section below).

Bundesen, 1990; Gobell, Tseng, & Sperling, 2004; Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980). Empirical work has also justified this connection, with rates converging with other measures of cognitive capacity (Donkin et al., 2014; Eidels, Donkin, Brown & Heathcote, 2010), and manipulations of capacity (e.g., adding a dual-task load) having the expected effects on rates (Castro, Strayer, Matzke & Heathcote, submitted; Logan et al., 2014).

The PMDC model allows for finer-grained analysis of capacity effects than do simple resource theories. For example, the model distinguishes the *quantity* of evidence accumulation from the *quality* of evidence accumulation. The quantity of accumulation is given by summing the rates for the matching and mismatching accumulators ('matching' refers to the accumulator for the response that matches the stimulus, i.e., the 'correct' response; 'mismatching' refers to the accumulator for the response that does not match the stimulus, i.e., the 'incorrect' response), whereas the quality of accumulation is given by the difference between the rates for the matching and mismatching accumulators. This distinction between the quantity and quality of accumulation is not as apparent in the diffusion decision model (DDM; Ratcliff, 1978), which has been a commonly used measure of PM cost (e.g., Horn & Bayen, 2015).²

As reviewed, most PM theories assume that PM cost to non-PM trials results from reduced ongoing task capacity, the idea being that individuals need to orient attention towards monitoring for PM in case they are presented a PM item (e.g., Einstein & McDaniel, 2005; Smith, 2003). Testing this idea requires comparing ongoing task performance with and without PM demand (i.e., comparing accumulation rates to non-PM items in PM blocks with

² Ratcliff, Voskuilen, and Teodorescu (2018) have recently suggested that similar distinctions may be made with the DDM by jointly examining mean accumulation rates and rate variability.

accumulation rates to non-PM items in control blocks). The majority of studies modeling PM costs have reported no change in either the quality or quantity of evidence accumulation to non-PM items across PM and control blocks, suggesting no change in the allocation or availability of resources as a result of PM (Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017, 2018). In contrast, one recent study did find some evidence of reduced processing quality under PM load (Anderson et al., 2018). However, this experiment did not counterbalance PM and control conditions, so it is possible that the evidence of reduced capacity was due to order effects rather than PM. Further contributing to the ambiguity of this effect, it only appeared in the diffusion decision model, and not in the LBA which provided a better fit to the data.

Prospective Memory in Complex Dynamic Tasks

To summarize the reviewed findings, it appears that in basic laboratory paradigms, individuals primarily manage PM demands by exerting cognitive control over ongoing task and PM processes, and that the capacity of the ongoing task is not affected by PM demands. However, the previous data does not imply that PM monitoring never affects ongoing task resources. Indeed, a benchmark of resource theory is that tasks drawing on the same pool of resources may not interfere with each other unless the total demand for resources exceeds the capacity limit of the cognitive system (Navon & Gopher, 1979; Norman & Bobrow, 1976). As such, basic laboratory tasks may leave resources idle, meaning that there is sufficient capacity available for monitoring in PM blocks without diverting resources from the ongoing task or increasing the amount of available resources (see Rummel, Smeekens, & Kane; 2017). This may render basic paradigms unable to detect capacity effects of PM that emerge when greater demands are placed on the cognitive system.

Air traffic control is a prototypical example of a job in which people need to manage the demands of PM in a complex and dynamic environment (see Dismukes, 2010, 2012; Loft, 2014). Air traffic controllers are frequently faced with the requirement to remember to perform deferred actions. For example, a controller might need to remember to provide information to an aircraft when it passes a waypoint or reaches its top of descent. Alternatively, a controller might need to remember to change the route or level of one aircraft after giving a clearance to another aircraft, or to put an aircraft into a holding pattern. Controllers often do this when they are under significant time pressure (Loft et al., 2007). For these reasons there is a growing body of both laboratory and field research examining PM using air traffic control tasks. These studies have shown that PM produces numerous performance costs, including slower acceptance and hand-off of aircraft, slower detection of conflicts between aircraft, and increased rates of missed conflicts (Loft, Chapman, & Smith, 2016; Loft, Finnerty, & Remington, 2011; Loft, Percy, & Remington, 2015; Loft, Smith, & Remington, 2013; Loft & Remington, 2010; Loft, Smith, & Bhaskara, 2011; Loukopoulos, Dismukes, & Barshi, 2009). These costs observed by Loft and colleagues, which affected both accuracy and RT, appear more consistent with resource sharing than increased thresholds, inconsistent with the results from basic PM paradigms in which raised ongoing task thresholds typically allow participants to maintain similar ongoing task accuracy in control and PM blocks (e.g., Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017, 2018).

In the current study, we apply the PMDC model so that we can measure capacity and control as has been done in the basic research. In our study, participants make decisions about whether two aircraft will come into conflict at some point in the future. In some blocks of trials, in addition to detecting conflicts, participants are required to press an alternative PM response

key instead of a conflict or non-conflict ongoing task response for aircraft with a callsign containing two of the same letter (e.g., APA169, RTR451). Comparing model parameters between control and PM blocks allows us to examine how PM affects attentional resources (e.g., capacity-sharing) and cognitive control mechanisms. To further test the role of capacity sharing in PM cost, we introduce a further *within-subjects* manipulation of ongoing task demands: time pressure. This allows us to examine whether capacity for PM and ongoing processes trades off across different levels of time pressure, which is important because such trade-offs are a critical indicator of capacity sharing (Navon & Gopher, 1979). To impose time pressure on responses, we manipulated the number of items that needed to be sequentially responded to (trial load), and the total time available to make that set of responses (trial duration). This was done to check whether trial load and time available both induce quantitatively similar time pressure effects, since this is not always the case in air traffic control or similar complex dynamic tasks (Loft et al., 2007; also see Hendy, Liao, & Milgram, 1997; Palada, Neal, Tay, & Heathcote, 2018). We now discuss testing the PMDC model as it would apply to the air traffic control task.

Testing Capacity and Cognitive Control

One question of interest is whether PMDC is capable of fitting the performance data from our conflict detection task. Evidence accumulation models have proved useful for understanding conflict detection in the past (Neal & Kwantes, 2009; Vuckovic, Kwantes, Humphreys, & Neal, 2014; Vuckovic, Kwantes, & Neal, 2013), accounting for accuracy and RT, and providing sensible psychological interpretations of the effects of manipulations of bias and speed-accuracy instructions. However, previous models required task-specific inputs such as relative speed and angle of approach, and thus have limited generalizability beyond the specific scenarios on which they are trained. Our goal here is a more general model, which requires fitting the model to each

individual's RT distributions and observed responses simultaneously. For the most part, evidence accumulation models have only been fit in this way to short time scale decisions (typically < 1 second), but recent studies have revealed promise for fitting to longer time scale decisions. For example, Lerche and Voss (2017) demonstrated good fits of the diffusion decision model to decisions with mean RTs over 7 seconds. Further, they found that tests of selective influence held, that is, manipulations affected rates where they were expected to (e.g., stimulus quality increased rates), and thresholds where they were expected to (e.g., strategy influenced thresholds). Palada et al. (2016; 2018) also found that evidence accumulation models (DDM and LBA) could fit slow RTs, pass tests of selective influence, and measure various forms of cognitive capacity, when fitting performance to a complex dynamic unmanned aerial vehicle surveillance task. Thus, we have reasonable grounds to expect that PMDC may fit to our data. Providing it can, the way in which it does so can give insight into the latent cognitive control and resource allocation mechanisms underlying the data, as we discuss further below.

Proactive control. We expect that participants will use cognitive control to manage demands associated with time pressure, in line with previous modeling of the speed-accuracy trade off (e.g., Dutilh, Wagenmakers, Visser, & van der Maas, 2011; Forstmann et al., 2011; Usher, Olami, & McClelland, 2002). For example, participants may adjust their ongoing task thresholds to avoid hitting trial deadlines under high pressure (i.e., when there is little time available per decision), or to take advantage of lax trial deadlines under low pressure (i.e., when there is more time available per decision; Ratcliff & Rouder, 1998). Raising thresholds allows one to gather more evidence before responding, which results in slower but more accurate decisions. Lowering thresholds allows one to spend less time gathering evidence and make faster responses at the expense of accuracy. Participants may also adapt to time pressure via control

over response bias (i.e., prioritizing one response over another). For example, in air traffic control, failing to detect a conflict has far greater safety implications than erroneously classifying a non-conflict as a conflict. As such, expert controllers strategically shift bias towards making conflict responses under time pressure to ensure aircraft remain separated (Loft, Bolland, Humphreys & Neal, 2009). In terms of the model, response bias is reflected in differences in thresholds between competing accumulators, whereby responding is biased in favour of the accumulator with the lower threshold.

We also expect to replicate the consistent findings from basic paradigms that individuals manage PM task demands by exerting proactive cognitive control over ongoing task thresholds (e.g., Heathcote et al., 2015; Strickland et al., 2017, 2018). That is, we expect to find higher ongoing task thresholds in PM blocks than in control blocks. Our relatively high PM frequency (see further below) may encourage stronger proactive control of ongoing task thresholds than in basic paradigms, as may the highly non-focal nature of the PM task. Further, if PM demands do decrease ongoing task capacity (also discussed further below), caution may also increase with PM to reduce ongoing task errors (as has been recently found with dual-task capacity costs; Castro, Strayer, Matzke & Heathcote, submitted).

The demands associated with time pressure and PM load may interact, leading different cognitive control strategies to trade off. For example, when time pressure is low, participants have little reason not to increase their ongoing task thresholds in PM blocks, whereas when time pressure is high, increasing ongoing task thresholds too much could lead to failing to perform responses before the deadline (Palada et al., 2018). Thus, ongoing task thresholds in PM blocks should increase more (relative to control blocks) at lower time pressures compared to higher time

pressures. Participants may also lower PM thresholds with increased time pressure to ensure that a PM response, if appropriate, is made before the response deadline.

Capacity Sharing. As with previous work, we examine how task demands affect the availability and allocation of resources by comparing ongoing task accumulation rates between different experimental conditions. We test for differences in resource availability in terms of the quantity (sum of matching and mismatching) of accumulation rates. Quantity reflects the total amount of resources deployed by the cognitive system, which might increase and decrease to meet task demands. For example, the quantity of available resources may increase as time pressure increases. PM load may place demands on the cognitive system similar to time pressure, and thus have similar effects on resource availability. In particular, quantity may increase in PM blocks if PM load increases the resources needed to perform the task, and perhaps also to compensate for PM-induced proactive control over ongoing task thresholds, which makes it difficult to respond before trial deadlines. Moreover, ongoing task and PM rates may trade-off across different levels of time pressure (i.e., capacity may be shunted from PM to ongoing processes as demands increase). Such a trade-off would demonstrate capacity sharing.

We also compare the quality of evidence accumulation (matching – mismatching accumulator ongoing task rates) across PM load and time pressure conditions. Palada et al. (2018) found that time pressure could cause a ‘redline’ to be crossed, after which performance rapidly degraded. Imposing shorter trial deadlines led to reduced processing quality, which participants attempted to compensate for by increasing their rate of information processing (i.e., quantity). This suggests that higher time pressure blocks should be more demanding than lower time pressure blocks, leading participants to divert resources from lower priority tasks to higher priority tasks when under heightened time pressure. PM load may also affect the quality of

ongoing task processing. Lower quality processing for ongoing task decisions under PM load would indicate that resources are being repurposed from the ongoing task to the PM task. Capacity effects may also occur because the PM cues (callsign) in our task are visually separate from the stimulus cues relevant to the ongoing conflict detection task (circular icons representing aircraft, speed indicator, relative distance of aircraft from the intersection), meaning that detecting callsign cues will likely require participants to orient attention away from the visual location of the primary task, towards the PM cue.

In addition to our behavioural measures of quantity and quality, we include subjective measures of effort, mental-, and temporal-demand (three subscales of the NASA-TLX; Hart & Staveland, 1988). This will let us check whether participants' subjective experiences of PM demands and time pressure align with the resource availability and resource allocation mechanisms in our model.

Reactive control. Trivially, we expect 'reactive excitation', in which PM accumulation rates are higher to PM items than non-PM items. In addition, previous applications of PMDC have found inhibition of ongoing task accumulation rates to PM items. That is, accumulation rates for conflict and non-conflict decisions should be lower for PM items, as compared with non-PM items from PM blocks. Because reactive control only affects responses to ongoing task stimuli when a PM item is present, it does not slow down overall responding very much, and so decreasing reactive control with increased time pressure is not likely to much improve the probability of meeting response deadlines. Thus, we do not have a strong reason to expect different levels of reactive control at different time pressures.

Method

Participants

Of 49 participants two were excluded (see Results), with 47 participants remaining (31 females). Participants were recruited from the UWA undergraduate research pool and had no prior experience with the air traffic control task. Ages ranged from 18 to 62 years ($M = 25.19$, $SD = 9.99$). Participants completed one two-hour testing session. All procedures were approved by the UWA Human Research Ethics Office.

Air Traffic Control Conflict Detection Task

The conflict detection task was designed by Fothergil, Loft and Neal (2009) in consultation with subject matter experts using principles of representative design in order to balance the competing demands of task fidelity, generality, and experimental control. The task has been previously used to develop and test a performance theory and computational model of expert conflict detection in air traffic control (Loft et al., 2009) and study PM (Loft, Smith, & Bhaskara, 2011). As illustrated in Figure 3, each trial of the conflict detection task presented a single pair of aircraft cruising at identical altitudes and converging on a common intersection in a fictitious en-route sector. The total area of the airspace was 180 nm (nautical miles) by 112.5 nm. Each aircraft had a data block that displayed the callsign, the aircraft type, the flight level, and the speed in knots (nautical miles per hour). Aircraft appeared within a circular air traffic control sector with a neutral grey background and flew straight paths (indicated by black lines) which converged at a 90-degree angle in the centre of the display. Aircraft position was updated every 20 ms. Participants had no control over the flight levels, velocities, or headings of the aircraft. They were required to classify each pair of aircraft as either 'conflict' or 'non-conflict' depending

on whether the aircraft would violate a 5 nm minimum separation distance at some point during their flight. On some trials one of the two aircraft would also contain a PM target feature which required execution of a PM response instead of the conflict or non-conflict ongoing task response. Within each trial, pairs of aircraft were presented sequentially; each pair disappearing from the screen once a response was made. As shown in Figure 3, the display included a countdown timer indicating the seconds remaining in each trial. Trials ended when the timer reached zero, after which any remaining aircraft not responded to would disappear and be recorded as non-responses. A 10 nm by 20 nm (approximately 2 cm by 4 cm on screen) scale marker, used as a reference for judging relative aircraft distance, was fixed on the left side of the display. Each aircraft had a probe vector line which showed the aircraft's heading and predicted position one minute into the future.

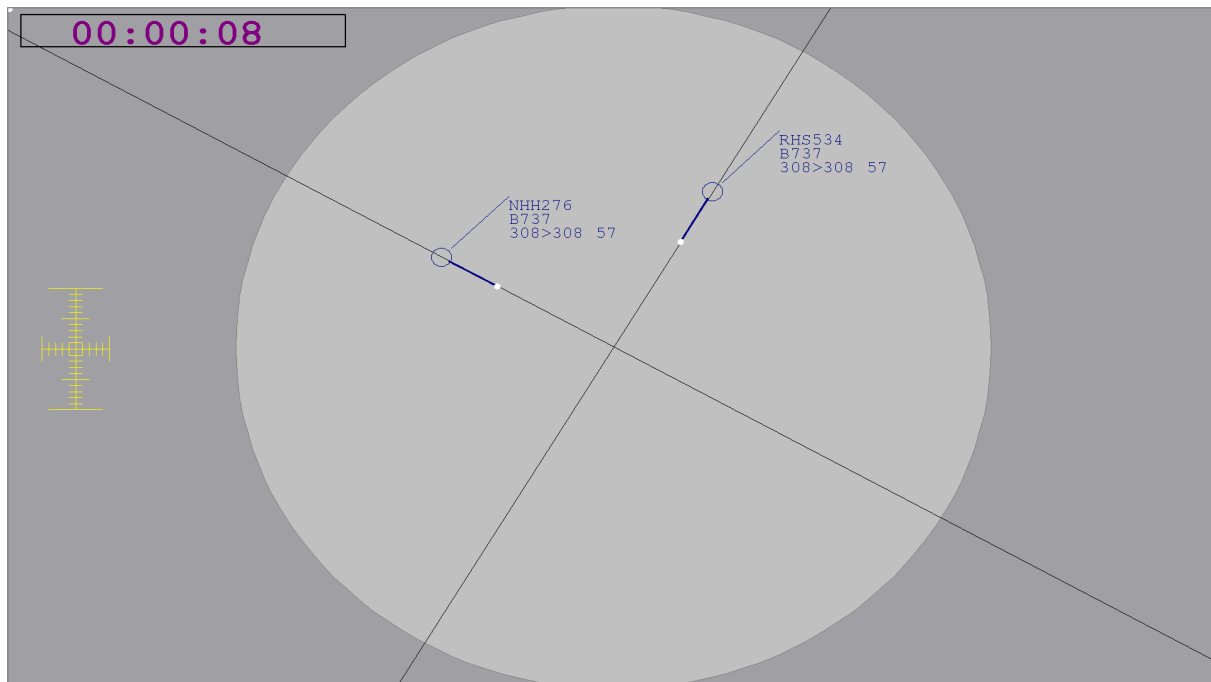


Figure 3. Air traffic control simulator display. Information blocks next to each aircraft show callsign (e.g., RHS534), aircraft type (B737), current and cleared altitude (e.g., 308>308), and airspeed (e.g., 57). Note that the airspeed indicator omits the ending zero of the true speed (e.g., '57' actually stands for 570 knots).

Experimental Stimuli and Design

Table 1 specifies the range of values for the features of aircraft pairs and the distribution they were drawn from. The angle of approach between aircraft was fixed at 90 degrees to avoid interactions between angle and perceived conflict status (e.g., Loft et al, 2009; Vuckovic et al, 2013). The flight level for all aircraft was fixed at 37,000 feet. To create the different conflict and non-conflict stimuli, each aircraft pair was assigned a miss distance (d_{min} : the distance between the aircraft at their point of closest approach) either less than or greater than the 5 nm separation standard. For conflict stimuli, d_{min} values were drawn from the uniform distribution [0,3] nm. For non-conflict stimuli, d_{min} values were drawn from the uniform distribution [7,10] nm. Because our primary interest here was in stable conflict detection performance, free from learning effects, we allowed other aircraft features to vary randomly. These features were speed, direction of approach, time to minimum separation (t_{min}), and order of passing at the crossing (i.e., faster aircraft first versus slower aircraft first). This ensured participants could not learn to use these features as predictive cues (Bowden & Loft, 2016; Loft, Neal, & Humphreys, 2007; Loft, Humphreys, & Neal, 2004).

Table 1

Range of spatial variables of aircraft stimuli

Spatial variable	Lower	Upper	Units
d_{min} (Conflicts)	0	3	nm
d_{min} (Non-conflicts)	7	10	nm
Airspeed	400	700	knots
Direction of approach	0	360	degrees
t_{min}	120	210	seconds
Order of passing	0	1	0 = fastest first, 1 = slowest first

Aircraft with callsigns containing two of the same letter (e.g., APA169, RTR451) were PM targets, thereby emulating the general task demand that operators monitoring perceptually-demanding displays can face to remember to perform a deferred task action in the future when they observe a particular event (Loft, 2014). The PM target is non-focal to conflict detection (Einstein & McDaniel, 2005), meaning that the evidence required to make PM decisions (i.e., assess aircraft callsign) is not required to make ongoing task conflict/non-conflict decisions (e.g., which requires assessing airspeed, relative distance, and position). On PM target trials only one of the aircraft on screen ever contained a PM target, never both. Participants were instructed to respond to PM targets by pressing an alternate PM key (e.g., ‘j’ or ‘d’) instead of the typical ongoing task (conflict/non-conflict) keys.

As illustrated in Table 2, participants performed four sets of trials, each containing a block of control trials and a block of PM trials. Block order (control- or PM-first) was counterbalanced between participants. In control blocks, participants were presented with a randomized sequence of 80 aircraft pairs (40 conflict and 40 non-conflict), with no PM targets. In PM blocks, participants were presented with a randomized sequence of 240 aircraft pairs (120 conflict and 120 non-conflict). Of these, a random 48 (24 conflict and 24 non-conflict) contained a PM target. Thus 20% (48/240) of PM block stimuli were PM targets.

Table 2

Details of experimental blocks with number of control and PM stimuli presented

Trial load (Decisions per trial)	Trial duration (s) (Overall time available)	Time pressure (s) (Average time available per decision)	Control block trials	PM block trials
2	12	6	80	240
2	8	4	80	240
5	20	4	80	240
5	10	2	80	240

To create time pressure (average available time per conflict detection decision), we manipulated trial load (decisions per trial) and trial duration (overall time available to respond to all decisions within a trial). Each trial had a load of either 2 or 5 (i.e., 2 or 5 aircraft pairs presented) with an associated trial duration (see Table 2). This resulted in 4 unique trial-load by trial-duration combinations; one with 6s per decision, two blocks with 4s per decision and one block with 2s per decision. Presentation order was counterbalanced across participants. We intentionally did not orthogonally cross trial load with trial duration. This was done to ensure trial load and trial duration remained at a reasonably engaging level of demand while also not

becoming impossibly difficult. Participants were informed of the trial load and trial duration prior to each block, giving them opportunity to adjust their strategies accordingly.

Procedure

Our primary goal when designing this experiment was to reliably model both ongoing task and PM responses in a more complex and dynamic task than is typically used in PM research. To this end, our task deviated slightly from traditional PM paradigms. Most PM studies conducted using the Einstein and McDaniel (1990) paradigm present only a small number of PM targets, and as such they do not produce enough data to reliably constrain a model of PM processes. Moreover, because the conflict detection task involves much slower RTs than typical laboratory paradigms, we were limited in the number of trials we could present to participants in a single 2-hour testing session. As such, we modified the paradigm by increasing the ratio of PM target trials to ongoing task trials to 1 in 5 trials (i.e., 20% of PM block stimuli were PM targets). This gives us more PM trials, which serve to reliably constrain our model, and increase the accuracy and precision of model fitting.

Each testing session consisted of a training phase and a test phase with total duration of 2 hours. During training participants received verbal task instructions, watched an on-screen demonstration, and completed a block of 40 training trials which included feedback after each response. PM targets were not included in training. During the experimental phase participants completed eight blocks of experimental trials, which did not include feedback.

Participants responded to each aircraft pair by pressing the *conflict* key or the *non-conflict* key. Participants were told that each aircraft pair would be presented sequentially (i.e., only two aircraft would appear on screen at a time), that all aircraft would be moving towards each other on converging flight paths which crossover in the centre of the display, and that a

number of spatial properties of the aircraft would vary from trial to trial, including their starting distance from the central crossing point, relative speed, and miss distance. Before each block of trials, participants saw visual instructions reminding them of the trial load and trial duration for that block. Depending on the block, participants received either control or PM instructions.

Before control blocks, participants were instructed that they only needed to make conflict and non-conflict responses. Before PM blocks, participants were instructed to press a PM response key instead of the conflict or non-conflict keys when they detected a PM target. Participants then completed a short distractor task and saw a final reminder to respond as quickly and accurately as possible before commencing the block.

Four response key assignments were counterbalanced across participants; 1) s = conflict, d = non-conflict, j = PM, 2) d = conflict, s = non-conflict, j = PM, 3) k = conflict, j = non-conflict, d = PM, and 4) j = conflict, k = non-conflict, d = PM. So that we could assume equal motor response time in our modeling (see Voss, Voss, & Klauer, 2010), participants were instructed to rest their fingers on their particular response key combination throughout the task. A screen with the text 'Press [Space] to continue' preceded each trial, and each trial began once the space-bar was pressed.

Trials ended when either all aircraft pairs had been responded to (2 pairs during blocks of low-load trials; 5 pairs during blocks of high-load trials) or when the response deadline expired (i.e., the timer reached zero). Aircraft pairs not responded to within the response deadline were recorded as non-responses. Participants were informed that conflict detection misses (responding non-conflict to conflict), conflict detection false alarms (responding conflict to non-conflict), PM misses (failing to make a PM response to a PM target), PM false alarms (making a PM response to a non-target), and non-responses, would be penalized equally. Aside from the training trials,

no further feedback was given concerning task performance. Participants took self-paced breaks between each block of trials and were also permitted short breaks at any point between trials if required. Subjective task demand after each block was assessed using the NASA Task Load Index (Hart & Staveland, 1988). The NASA-TLX comprises six items: three that tap task demand (i.e., mental, physical, and temporal demand) and three which tap the individual's subjective perception of exertion and task performance (i.e., performance, effort, and frustration). Each item is rated on a 21-point numerical scale, with higher scores being indicative of higher workload.

Results

Conventional statistical analyses are reported first in order to check whether our experimental manipulations had the expected effects on RT, accuracy, and non-response (miss) rate. Data from two participants was excluded; one participant who failed to complete all experimental blocks and one who made no PM responses at all. We excluded trials with outlying RTs, defined as less than 0.2s or 3 times the inter-quartile range / 1.349 (a robust measure of standard deviation) above the mean (1.52% of responses overall). Outliers were censored separately for each different time pressure condition. Overall, 6.3% of the data comprised non-response misses, ranging from 2.5% in the lowest time pressure condition, to 14.8% in the highest. Two kinds of extremely rare responses – incorrect ongoing task responses to PM items (1.5% of PM block responses), and PM responses to control-block ongoing-task stimuli (0.4% of all responses) – are not analysed further. The conventional statistical analyses compare mean accuracy and RT by stimulus type (conflict, non-conflict, PM) PM block (control, PM) and time pressure. Because trial load and trial duration were not crossed orthogonally, time pressure is

compared separately for each level of trial load. That is, at low trial load (2 decisions per trial) we compare time pressures of 6 and 4 seconds per decision, and at high trial load (5 decisions per trial) we compare time pressures of 4 and 2 seconds per decision.

In our significance testing for accuracy effects we used generalized linear mixed models with a probit link function. In our significance testing for mean correct RTs we used general linear mixed models.³ Analyses were conducted using the R package *lme4* (Bates, Machler, Bolker, & Walker, 2015), and significance assessed with Wald's chi-square tests (Fox & Weisberg, 2011), using an alpha level of 0.05. Post hoc tests applied Bonferroni's correction for alpha inflation. The results of our analyses are tabulated in the supplementary materials (Tables S1-S4). All standard errors reported in text and displayed in graphs were calculated using the within-subject bias-corrected method (Morey, 2008).

Conflict Detection (Non-PM) Trials

Accuracy was lower for conflicts (67.4%) compared with non-conflicts (80.6%) and slightly lower under PM load compared with control (Control: $M = 74.9\%$, $SE = 3.2\%$; PM: $M = 73.1\%$, $SE = 3.3\%$). Conflict detection accuracy decreased as time pressure increased, under both low trial-load (6s: $M = 76.9\%$, $SE = 2.7\%$; 4s: $M = 74.8\%$, $SE = 2.7\%$) $t = 2.60$, $df = 46$, $p = .012$,

³ We used two methods of analysis to check that the inferences derived from each method agreed. For our accuracy analyses, we compared a binomial model with a probit link function on accuracy to a binary logistic regression model. The inferences derived from both methods were the same. For our RT analyses, we compared a Gaussian model on mean RT to a Gaussian model on the logarithm of raw RT, the latter having more statistical power than the former. The inferences derived from both were the same. The log(RT) method revealed two interactions (stimulus by PM block, and time pressure by PM block) on RT that were not present in the less powerful mean RT model, but because these effects were small and not of theoretical importance we report the mean RT analysis. Visual inspection of residual plots revealed that residuals were approximately normally distributed for accuracy, mean RT, and log(RT) models.

Cohen's d = 0.38, and high trial-load conditions (4s: $M = 75.7\%$, $SE = 2.5\%$; 2s: $M = 68.6\%$, $SE = 2.6\%$) $t = 7.82$, $df = 46$, $p < .001$, $d = 1.14$. Conflict detection accuracy was not significantly affected by trial load when comparing cells with equal time pressure. There was a significant interaction between PM block and time pressure on accuracy, such that the cost to PM block accuracy was greatest when trial load and time pressure were also high (Control: $M = 70.9\%$, $SE = 2.1\%$; PM: $M = 66.3\%$, $SE = 2.5\%$), $t = 4.25$, $df = 46$, $p < .001$, $d = 0.62$.

Mean RT was slower for conflicts (3.09s) compared with non-conflicts (2.75s), slower for errors (2.99s) compared with correct responses (2.91s), and slower during PM blocks than control blocks (Control: $M = 2.65s$, $SE = 0.17s$; PM: $M = 3.07s$, $SE = 0.17s$). Mean RTs were significantly faster under higher time pressure for both low trial-load (6s: $M = 3.65s$, $SE = 0.14s$; 4s: $M = 2.80s$, $SE = 0.08s$) $t = 7.14$, $df = 46$, $p < .001$, $d = 1.04$, and high trial-load conditions (4s: $M = 2.95s$, $SE = 0.10s$; 2s: $M = 2.04s$, $SE = 0.07s$) $t = 12.26$, $df = 46$, $p < .001$, $d = 1.79$. Mean RTs were also slower under high trial load compared with low trial load when comparing cells with equal time pressure, $t = -2.17$, $df = 46$, $p = .035$, $d = 0.32$, although this effect was small. There was no significant interaction between PM block and time pressure on mean RT.

To summarize, the addition of PM load resulted in slower (*Mean Difference* = 0.42s) and slightly less accurate (*Mean Difference* = 1.8%) ongoing conflict detection task performance, while increased time pressure led to faster (*Mean Difference* = 0.88s) but less accurate (*Mean Difference* = 4.6%) ongoing conflict detection task performance.

PM Trials

PM responses were scored correct if the participant made a PM response instead of an ongoing task (conflict/non-conflict) response on PM target trials. PM accuracy decreased as time pressure increased during both low trial-load (6s: $M = 81.7\%$, $SE = 1.8\%$; 4s: $M = 74.1\%$, $SE =$

2.1%) $t = 3.19$, $df = 46$, $p = .003$, $d = 0.47$, and high trial-load conditions (4s: $M = 73.8\%$, $SE = 1.9\%$; 2s: $M = 58.6\%$, $SE = 2.7\%$) $t = 4.94$, $df = 46$, $p < .001$, $d = 0.72$. PM accuracy was not significantly affected by trial load when comparing cells with equal time pressure. PM accuracy was weakly positively correlated with ongoing task accuracy ($r = .21$, $p < .001$).⁴

Mean RT was slower for PM errors (2.73s) compared with correct PM responses (1.77s) (in PM blocks). Mean RT for correct PM responses was significantly faster at higher levels of time pressure during both low trial-load (6s: $M = 2.00s$, $SE = 0.05s$; 4s: $M = 1.78s$, $SE = 0.05s$) $t = 3.96$, $df = 43$, $p < .001$, $d = 0.60$, and high trial-load conditions (4s: $M = 1.89s$, $SE = 0.05s$; 2s: $M = 1.58s$, $SE = 0.05s$) $t = 6.40$, $df = 44$, $p < .001$, $d = 0.95$. PM RTs were slower under high trial load compared with low trial load when comparing cells with equal time pressure, $t = -2.13$, $df = 43$, $p = .039$, $d = 0.32$, although this was a relatively small effect. There were no significant differences in PM accuracy or PM RT between conflict PM targets and non-conflict PM targets. PM false alarms (i.e., PM responses to non-PM stimuli in PM blocks) occurred on 0.56% of PM block trials with a mean RT of 2.45s. Mean PM RT was moderately positively correlated with mean ongoing task RT ($r = .48$, $p < .001$). To summarize, as with the ongoing task, increased time pressure led to faster (*Mean Difference* = 0.27s) but less accurate PM performance (*Mean Difference* = 11.4%).

⁴ The correlation between ongoing task and PM accuracy is consistent with the strong role of proactive control (discussed below) over ongoing task thresholds in PM blocks in this study (i.e., higher thresholds lead to both higher ongoing task accuracy and higher PM accuracy). The correlation between ongoing task and PM RT is likely due to individual differences in response caution, or statistical facilitation (i.e., slower PM accumulators win more frequently against slower ongoing task accumulators).

Non-responses

We ran a linear mixed effects model (Table S6 in the supplementary materials) to examine the effects of PM block and time pressure on non-response (miss) proportions (non-responses included non-responses to PM and ongoing task stimuli). Non-responses were slightly more frequent in PM blocks compared with control (Control: $M = 5.8\%$, $SE = 1.6\%$; PM: $M = 5.9\%$, $SE = 1.8\%$), and became more frequent as time pressure increased during both low trial-load (6s: $M = 2.2\%$, $SE = 1.0\%$; 4s: $M = 4.4\%$, $SE = 0.9\%$) $t = 2.71$, $df = 46$, $p = .009$, $d = 0.4$, and high trial-load conditions (4s: $M = 3.0\%$, $SE = 0.8\%$; 2s: $M = 13.8\%$, $SE = 1.7\%$) $t = 6.29$, $df = 46$, $p < .001$, $d = 0.92$. The proportion of non-responses was not significantly affected by trial load when comparing cells with equal time pressure, $t = -1.89$, $df = 46$, $p = .07$, $d = 0.28$. There was a significant interaction between PM block and time pressure on non-responses, such that the increase in non-responses from control to PM blocks was greatest when trial load and time pressure were both high (Control: $M = 12.0\%$, $SE = 1.4\%$; PM: $M = 14.7\%$, $SE = 1.7\%$), $t = 2.17$, $df = 46$, $p = .04$, $d = 0.32$. This suggests a drawback of using the proactive control strategy under high time pressure. That is, higher ongoing task thresholds caused decisions to exceed the trial deadline more often, resulting in more non-responses to ongoing task stimuli.

NASA-TLX Demand Ratings

We ran a linear mixed effects model (Tables S7-S9 in the supplementary materials) to examine the effects of PM block and time pressure on self-report effort, mental-, and temporal-demand ratings. Effort was significantly higher in PM blocks than control blocks (Control: $M = 12.08$, $SE = 0.68$; PM: $M = 13.50$, $SE = 0.71$). Effort was also higher in high time pressure blocks during both low trial-load (6s: $M = 11.38$, $SE = 0.57$; 4s: $M = 13.13$, $SE = 0.52$) $t = 2.80$, $df = 46$, $p = .007$, $d = 0.41$, and high trial-load conditions (4s: $M = 11.48$, $SE = 0.51$; 2s: $M = 15.18$, $SE =$

0.46) $t = 6.93$, $df = 46$, $p < .001$, $d = 1.01$. Effort was lower during high trial load compared with low trial load when comparing cells with equal time pressure, $t = 2.93$, $df = 46$, $p = .005$, $d = 0.43$. There was no interaction between PM block and time pressure on effort.

Similarly, mental demand was significantly higher in PM blocks than control blocks (Control: $M = 12.31$, $SE = 0.57$; PM: $M = 14.60$, $SE = 0.58$). Mental demand was also higher at higher levels of time pressure during both low trial-load (6s: $M = 12.47$, $SE = 0.48$; 4s: $M = 13.41$, $SE = 0.39$) $t = 2.16$, $df = 46$, $p = .036$, $d = 0.31$, and high trial-load conditions (4s: $M = 12.35$, $SE = 0.44$; 2s: $M = 15.60$, $SE = 0.51$) $t = 6.55$, $df = 46$, $p < .001$, $d = 0.96$. Mental demand was lower during high trial load compared with low trial load when comparing cells with equal time pressure, $t = 2.78$, $df = 46$, $p = .008$, $d = 0.41$. There was no interaction between PM block and time pressure on mental demand.

Temporal demand was significantly higher in PM blocks than control blocks (Control: $M = 11.75$, $SE = 0.82$; PM: $M = 13.37$, $SE = 0.97$). Temporal demand was also higher at higher levels of time pressure during both low trial-load (6s: $M = 9.48$, $SE = 0.56$; 4s: $M = 13.15$, $SE = 0.56$) $t = 6.02$, $df = 46$, $p < .001$, $d = 0.88$, and high trial-load conditions (4s: $M = 10.61$, $SE = 0.61$; 2s: $M = 17.01$, $SE = 0.54$) $t = 9.80$, $df = 46$, $p < .001$, $d = 1.43$. Temporal demand was lower during high trial load compared with low trial load when comparing cells with equal time pressure, $t = 3.58$, $df = 46$, $p = .001$, $d = 0.52$. There was no interaction between PM block and time pressure on temporal demand. Taken together, the effort, mental-, and temporal-demand

findings give us confidence that our PM and time pressure manipulations were effective at inducing different levels of subjective task demand.⁵

Model Analysis

Our design includes several factors over which model parameters can vary, including latent response (i.e., conflict, non-conflict, and PM accumulators), and three manifest factors, stimulus type, time pressure/trial load, and PM demand. The latent response factor refers to the accumulators that can lead to each response (conflict, non-conflict, PM). The latent response factor corresponds to the accumulators, and not the response that was actually observed; the observed response is predicted by, not included in, the model. The stimulus type factor had four levels: *non-PM conflict*, *non-PM non-conflict*, *PM conflict*, and *PM non-conflict*. Since the trial-load (2 versus 5 decisions per trial) and task-duration factors were not crossed orthogonally, they were captured by a four-level composite factor created by manipulating trial load and trial duration with the following levels: 6s per decision (2 decisions per trial), 4s per decision (2 decisions per trial), 4s per decision (5 decisions per trial), 2s per decision (5 decisions per trial). The PM demand factor had two levels: *control* (i.e., no PM demand) and *PM*.

To reduce model complexity, we applied several theoretically sensible *a priori* constraints on which factors each parameter could vary. First, we estimated one common A parameter for all accumulators and conditions, as is common practice in LBA modeling. Second,

⁵ The correlations between the subjective effort ratings and various model parameters (e.g., rate quality and quantity) were small ($r = 0.1-0.2$) and mostly non-significant, but trended in sensible directions (i.e., rate quantity was positively correlated with effort, rate quality was negatively correlated with effort). The lack of statistical significance was likely due to the sample size and coarse nature of the self-report effort ratings.

we allowed the sv parameter to vary by stimulus and accumulator factors but not over PM block or time pressure. This is more flexible than most previous LBA modeling, which typically only allows sv to vary as a function of whether the latent accumulator matches or does not match the stimulus⁶ (but see Heathcote & Love, 2012; Osth, Bora, Dennis, & Heathcote, 2017; Strickland et al., 2018, for exceptions). We used this more flexible approach because in our model there are two types of 'correct' response for PM trials (i.e., correct PM and correct ongoing task decision). We fixed the sv parameter for PM false alarms (i.e., 'PM' responses to non-PM stimuli) at 0.5, because one accumulator parameter must be fixed to an arbitrary value as a scaling parameter (Donkin, Brown, & Heathcote, 2009).

Third, we estimated only one non-decision time (ter) parameter for each participant. This was done because our design minimized any potential differences in the motor movement required to make each response (i.e., participants kept their fingers positioned above the response key which were all located on one keyboard row). In addition, previous research has shown non-decision time does not play a role in PM cost for the LBA (e.g., Anderson et al., 2018; Heathcote et al., 2015; Strickland et al., 2017, 2018). Finally, due to very low numbers of PM false alarms (PM responses to non-PM stimuli in PM blocks) we pooled estimates of both accumulation rate and variance (v and sv) across all experiment factors to give one PM false alarm rate and one corresponding sv parameter (which was used as a fixed scaling parameter as mentioned above). These *a priori* restrictions resulted in an 89 parameter most flexible 'top' model with one A , one

⁶ We also fit a version of the model with only one sv parameter. This model had fewer parameters (75), produced visually similar fits, and either preserved or exaggerated the direction of all effects found in the more flexible model.

ter, 20 *B*, 57 *v*, and 10 *sv* parameters. We compared this flexible top model against several simpler, more constrained variants as outlined in the Model Selection section below.

Sampling

We estimated model parameters using the DMC software (Heathcote, et al., 2018) to perform Bayesian estimation, which results in probability distributions that reflect the certainty about parameter values. Although in principle we could have fit a hierarchical model to estimate the common population distributions of each parameter, we estimated parameters separately for each participant for several reasons. First, since this is the first time that such a model has been fit to this kind of task, we did not have adequate knowledge of the appropriate form of the population-level distributions. Because inappropriate population-level assumptions can introduce inappropriate shrinkage effects in hierarchical models, fitting to individual participants avoids such issues. Second, because of the large number of participants in our sample and the complexity of our models, hierarchical methods proved too computationally expensive to fit (estimated at several months of multi-core server time per fit).

Bayesian analysis requires that the researcher specify prior beliefs about the probabilities of parameters and the form of their distributions before observing the data. However, because of our large sample sizes and use of inference based on posterior probability distributions, the influence of our particular choice of priors on the final parameter estimates was negligible. Since these analysis techniques have not been used on a dynamic complex task, we did not have strong reasons to prefer one particular set of priors over others. We used relatively non-informative priors similar to Strickland et al. (2018), but with higher threshold priors to account for our slower mean RTs (Table 3). All prior values were the same over control and PM blocks and the different levels of time pressure.

Posterior parameter distributions were estimated using the differential evolution Markov-chain Monte-Carlo (DE-MCMC) algorithm (Turner, Sederberg, Brown, & Steyvers, 2013). DE-MCMC is more adept than conventional samplers at handling the high parameter correlations common to accumulate-to-threshold models. The number of chains was three times the number of parameters (e.g., for an 84-parameter model there were 252 chains per parameter). Chains were thinned by 20, meaning that one iteration in every 20 was kept. Sampling continued for each participant until a small Gelman's (2014) multivariate potential scale reduction factor (<1.1) indicated convergence, stationarity, and mixing. Convergence, stationarity, and mixing were verified by visual inspection. We retained the same number of samples for each participant: each of the 252 chains was 120 iterations long, producing 30,240 samples of each parameter's posterior distribution for each participant.

Table 3

Prior distributions

Parameter	Distribution	Mean	SD	Lower	Upper
A	Truncated normal	3	1	0	10
B	Truncated normal	2	1	0	None
v (Matching ongoing task response)	Truncated normal	1	2	None	None
v (Mismatching ongoing task response)	Truncated normal	0	2	None	None
v (PM hit)	Truncated normal	1	2	None	None
v (PM false alarm)	Truncated normal	0	2	None	None
sv	Truncated normal	0.5	1	0	None
t0	Uniform		1	0.1	1

Model Results

Model Fits: Accuracy and RT

To evaluate fit, we sampled 100 posterior predictions for each participant and then averaged over participants. The model provided good fits to both ongoing task and PM accuracy (Figures S1-S2 in the supplementary materials) and gave a good account of the entire distribution of response times (Figures S3-S5). The model provided a close fit to the differences in manifest accuracy and RT observed across PM and control blocks and across different time pressures.

Model Fits: Non-response Proportions

As noted earlier, our data contained significant differences in non-response proportions between time pressure blocks. Because non-responses were not included in model fitting, this gave us an opportunity to test PMDC's consistency with unseen data. That is, we assessed how well the model could make out-of-sample predictions of empirical non-response proportions. In order to do this, we simulated data out of the selected model (see Table 4) and matched the order of the simulated stimuli and responses to the actual presentation order experienced by each participant (this was done to ensure that the simulated trials had the same stimulus-response content as the empirical trials). Whenever the cumulative sum of simulated RTs within a trial exceeded that trial's deadline, a non-response was predicted. Using this method, 100 posterior predictions for non-response proportions were sampled for each participant and predictions were then averaged over all participants. We then compared the predicted non-responses with observed non-response proportions across the different levels of time pressure (i.e., different response deadlines) for both low and high trial-load conditions in control and PM blocks. Figure 4 shows observed versus predicted non-response proportions. Predicted non-response

proportions closely match the empirical non-response proportions, demonstrating the ability of the model to predict data out of sample.

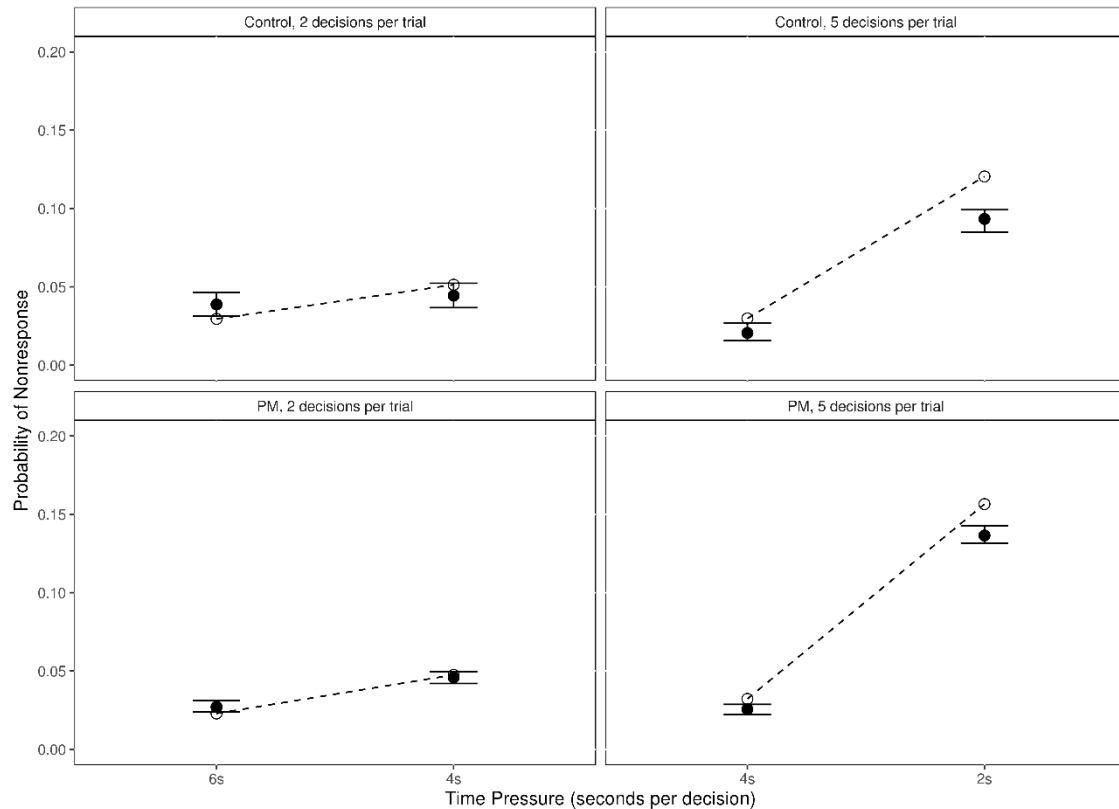


Figure 4. Model fits to non-response proportions by PM block and time pressure. Data effects are represented by unfilled circles. Model predictions are represented by filled circles with 95% credible intervals.

Model Selection

We applied model selection to assess whether we could justify constraining model parameters over blocked experimental conditions (e.g., PM block, time pressure) to obtain a simpler model with fewer parameters. To select between models, we used the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002), a measure which considers both goodness of fit and model complexity (number of parameters). In general,

models with smaller DIC values are to be preferred. Table 4 shows each model we compared, its number of parameters, and DIC value.

Table 4

DIC model selection. Lower DIC indicates more preference for the model.

Model	Parameters	DIC	DIC – minimum DIC
Top model	89	187068	976
Selected model	84	186092	0
Thresholds fixed over PM block	81	190690	4598
Rates fixed over PM block	73	188370	2278
Thresholds fixed over time pressure	74	190640	4548
Rates fixed over time pressure	47	190742	4650

Starting with the fully flexible top model, we built several simpler variants by systematically constraining ongoing task threshold and ongoing task rate parameters over PM and time pressure. This allowed us to assess whether it was necessary to vary ongoing task thresholds and/or rates to account for the observed PM demand and time pressure effects (i.e., to test whether proactive control and/or capacity-sharing were necessary mechanisms in the model). PM thresholds and rates were left free. We compared the following four constrained models to the top model: a model in which rates could vary across PM and control blocks but thresholds could not; a model in which thresholds could vary across PM and control blocks but rates could not; a model in which rates could vary by time pressure but thresholds could not; and a model in which thresholds could vary by time pressure but rates could not.

As Table 4 shows, in each case the simpler model was rejected in favour of the fully flexible top model, suggesting that it is necessary to allow both ongoing task threshold and ongoing task rate parameters to vary over PM and time pressure (i.e., both parameters are influenced by PM and time pressure manipulations and are important in explaining the observed data).

Finally, we tested an additional model (the selected model) that, like the top model, allowed both rates and thresholds to vary over both PM and time pressure but included a slight simplification: the PM rate parameter was constrained to not vary over stimulus type (i.e., PM conflicts and PM non-conflicts had the same rate). This simplification makes theoretical sense, since the evidence used to make a PM decision was independent of the evidence used to make either conflict or non-conflict ongoing task decisions. This slightly simpler model produced the smallest DIC value and was thus selected as our preferred model for further analysis.

The results of model selection suggest that both rates and thresholds have some role in explaining ongoing task and PM accuracy and RT under different levels of PM demand and time pressure. However, we wanted to further break down how each part of the model contributes to observed accuracy and RT. To this end, in the next section we test the direction and magnitude of differences between conditions in the parameters of the selected model. Testing the direction of effects is important because it allows us to identify how cognitive control and resource allocation mechanisms contribute to performance, and thus to distinguish between competing theories of PM and task demand, whose predictions are also directional (e.g., capacity-sharing theories predict lower rates in PM blocks than in control blocks). Testing the magnitude of effects is similarly important, as it indicates which processes contribute the most to a given effect

(such as costs to accuracy and RT) and which are most affected by a given experimental manipulation (such as time pressure).

Model Summary

To summarize the parameters of the group of participants, we created a subject-average posterior distribution. This was obtained by computing the mean of each posterior sample over all participants for each parameter. Our primary theoretical interest was in threshold and rate parameters for the ongoing and PM tasks, which we explore in detail in the following sections. The other parameters all had reasonable values. The non-decision time mean of the subject-average posterior distribution was 0.35 seconds (posterior $SD = 0.01$ seconds). The A posterior mean was 3.34 (posterior $SD = 0.04$). The sv posterior means and SDs are summarized in Table 5. Consistent with other LBA studies, sv parameters for the ongoing task are lower for matching response accumulators compared with mismatching response accumulators when it was a non-PM stimulus. For PM stimuli there was little difference in sv between accumulators. The accumulation rate for PM false alarms was -1.30 (posterior $SD = 0.05$).

Table 5

Mean (SD) of the average posterior sv parameter samples

Accumulator	Stimulus			
	Conflict	Non-conflict	PM (Conflict)	PM (Non-conflict)
Conflict	0.35 (0.01)	0.54 (0.01)	0.52 (0.02)	0.59 (0.03)
Non-conflict	0.53 (0.01)	0.46 (0.01)	0.53 (0.03)	0.63 (0.02)
PM	Fixed at 0.5		1.14 (0.03)	

The direction and magnitude of differences in ongoing and PM threshold and rate parameters for the selected model across conditions were examined to assess how well they correspond to the theoretical predictions of capacity sharing, proactive control, reactive control, and arousal/availability. We calculated posterior distributions of the differences between experimental conditions. For example, to test the difference between ongoing task response thresholds in control and PM blocks (i.e., testing the proactive control account of PM costs), we subtracted the control block threshold from the PM block threshold for every posterior sample, thus obtaining the posterior probability distribution of the difference between control and PM thresholds. Differences were calculated independently for each participant before being averaged across participants to create a subject-averaged posterior difference distribution. For each subject-averaged difference distribution we report a Bayesian *p*-value (Klauer, 2010), which indicates the one-tailed probability that the effect does not run in the most sampled direction.

Due to the large number of trials per participant in our design, almost all our observed parameter differences have *p*-values very close to zero, indicating a very high probability that an effect was present. However, some of our parameter differences were much larger in magnitude than others. As such, we illustrate the magnitude of the effect by reporting the standardized difference between parameters (i.e., M / SD of the posterior difference distribution). Because our posterior parameter distributions are approximately normal, this standardized statistic can be interpreted in a similar way to a *Z*-score. We therefore refer to this statistic as *Z* from here on. *Z*-score effect sizes and *p*-values for parameter comparisons are shown in Tables S10-S19 in the supplementary materials.

Proactive Control of Thresholds under Time Pressure and PM Demand

Figure 5 (top panel) compares ongoing task thresholds across control and PM blocks for each time pressure condition (averaged over accumulators). Consistent with proactive control, ongoing task thresholds were higher in PM than control blocks for both conflict ($Z = 35.80, p < .001$) and non-conflict ($Z = 44.32, p < .001$) accumulators. In both control and PM blocks, average ongoing task thresholds decreased as time pressure increased, under both low trial-load (Control: $Z = 3.09, p = .001$; PM: $Z = 7.16, p < .001$) and high trial-load (Control: $Z = 4.33, p < .001$; PM: $Z = 21.34, p < .001$) conditions. Average ongoing task thresholds were also slightly lower in high trial-load compared to low trial-load conditions when comparing cells with equal time pressure (Control: $Z = 2.84, p = .002$; PM: $Z = 3.33, p < .001$). These effects suggest that participants adapt to PM and time pressure demands by making *a priori* threshold adjustments, consistent with proactive control.

PM-induced proactive control of ongoing task thresholds interacted with time pressure. Specifically, the magnitude of ongoing task threshold adjustments between control and PM blocks got smaller as time pressure increased. As shown in the top panel of Figure 5, the average size of PM block-control block threshold differences decreased under high time pressure relative to low time pressure during both low trial-load ($Z = 2.67, p = .004$) and high trial-load ($Z = 12.38, p < .001$) conditions. The magnitude of PM block-control block ongoing task threshold differences was not significantly affected by trial load when comparing cells with equal time pressure ($Z = 0.25, p = .40$).

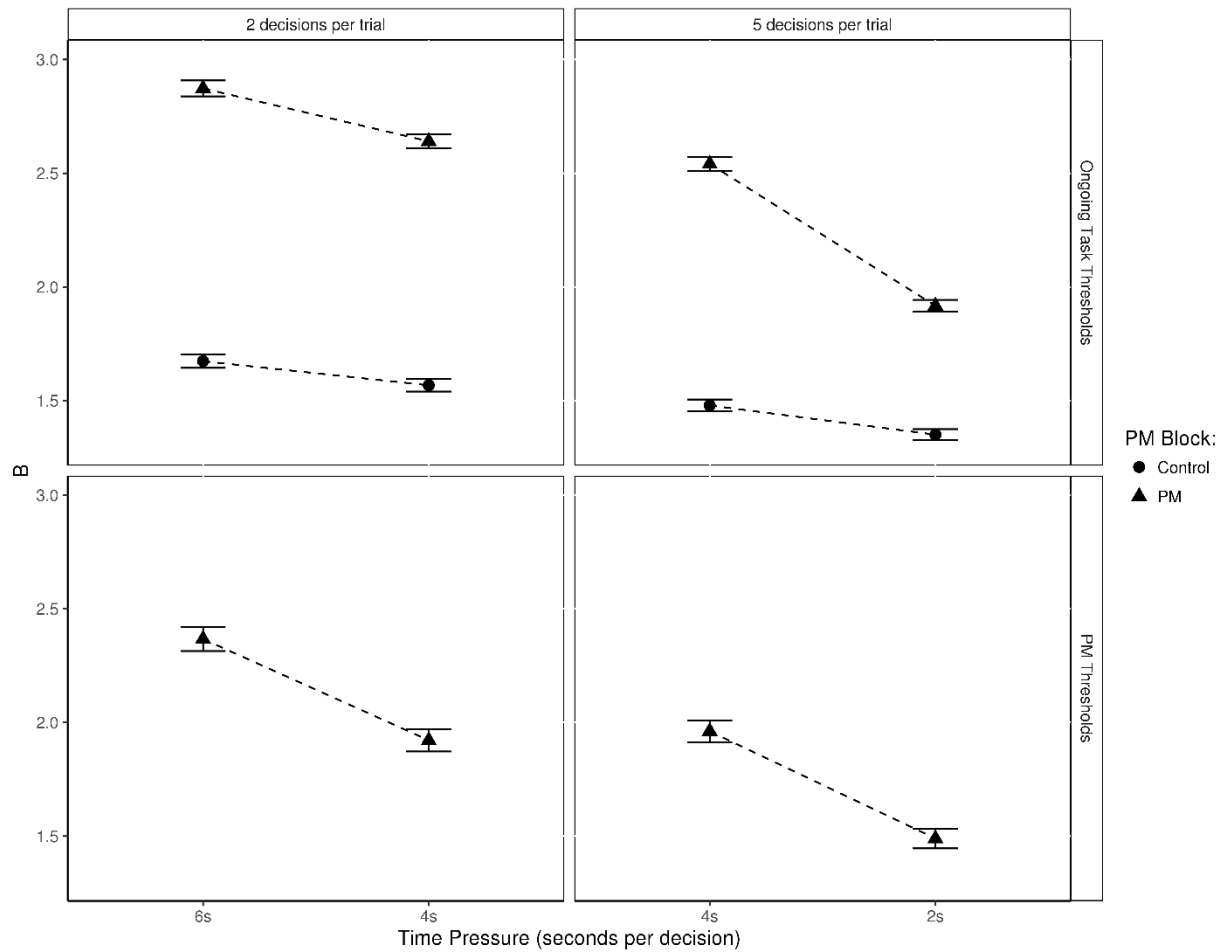


Figure 5. Average ongoing task and PM thresholds by time pressure and PM block. Central symbols represent posterior means. Error bars represent ± 1 posterior standard deviation.

Figure 5 (bottom panel) shows PM thresholds under different levels of time pressure for low trial-load (2 decisions per trial) and high trial-load (5 decisions per trial) conditions. PM thresholds decreased under high time pressure relative to low time pressure during both low trial-load ($Z = 7.98, p < .001$) and high trial-load ($Z = 8.64, p < .001$) conditions, but did not differ significantly by trial load when comparing cells with equal time pressure ($Z = -0.73, p = .23$). Overall, these shifts in PM thresholds with time pressure are consistent with prior modeling of

the speed-accuracy trade-off and support a proactive control account in which thresholds are strategically lowered to facilitate fast responding at the expense of accuracy.

Proactive Control of Response Bias under Time Pressure

Figure 6 plots response threshold bias (i.e., the difference between conflict and non-conflict thresholds) by time pressure. The dotted horizontal line represents the zero-bias point; the point where conflict and non-conflict thresholds are equal. Values above the zero-line indicate that non-conflict thresholds were lower than conflict thresholds (i.e., a bias toward responding 'non-conflict'). Likewise, values below the zero-line indicate that conflict thresholds were lower than non-conflict thresholds (i.e., a bias toward responding 'conflict'). Response bias tended to shift from favouring non-conflict responses under lower levels of time pressure to favouring conflict responses when time pressure is higher during PM blocks. This occurs under both low trial-load ($Z = 1.76, p = .04$) and high trial-load ($Z = 2.07, p = .02$) conditions. The same trend occurred (i.e., increased bias towards responding 'conflict') during control blocks but did not reach significance. Figure 6 shows that the same trend occurred from control blocks to PM blocks, suggesting the additional demands of the PM task also led to a shift in bias. Overall, this indicates that as task demands increase, participants tended to set lower conflict thresholds relative to non-conflict thresholds, although we note that these effects were relatively small. This strategy facilitates faster conflict responses and helps ensure that conflicts are not mis-categorised as non-conflicts but increases the frequency of conflict false alarms (i.e., 'conflict' responses to non-conflicts).

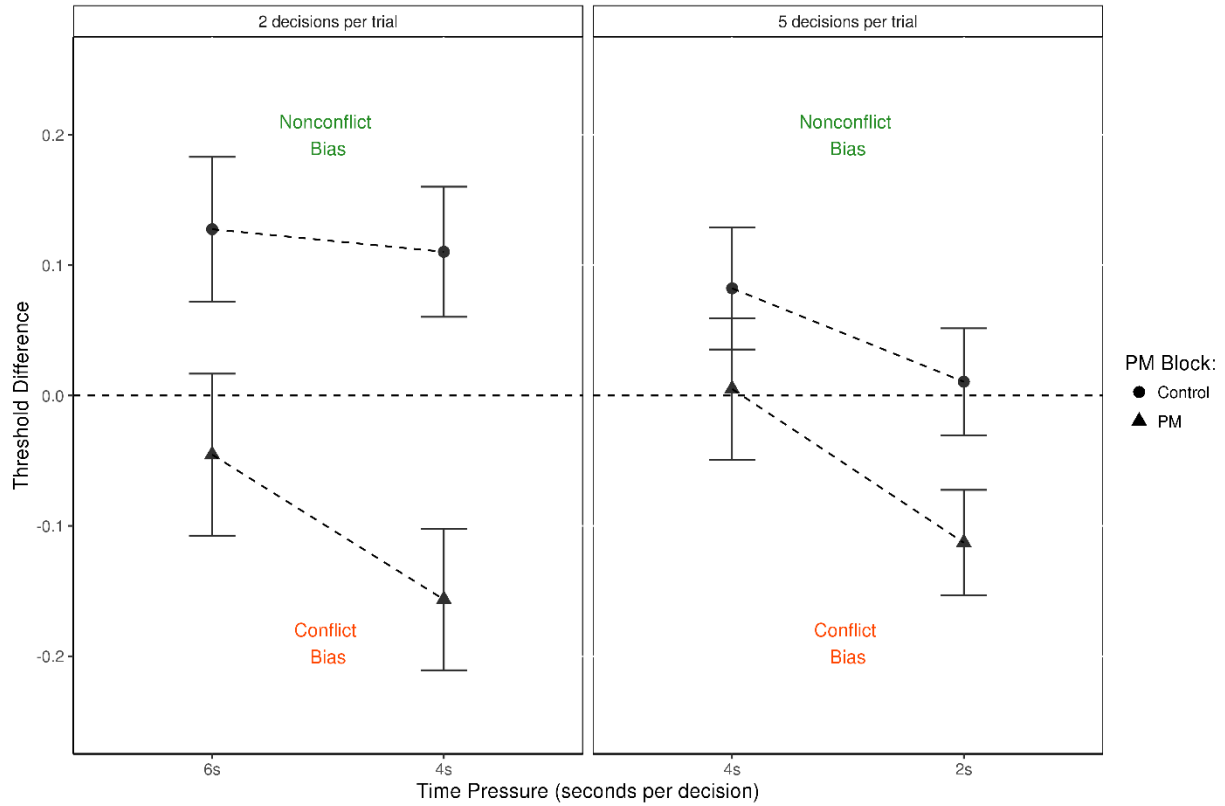


Figure 6. Ongoing task threshold bias by time pressure. Central symbols represent posterior mean difference between conflict and non-conflict thresholds. Error bars represent ± 1 posterior standard deviation. Dotted horizontal line represents the zero-bias point (i.e., equal thresholds). Above the zero-line indicates lower non-conflict than conflict thresholds. Below indicates lower conflict than non-conflict thresholds.

Capacity Sharing (Resource Quantity and Quality)

Quantity. As can be seen in Figure 7 (top panel), ongoing task rates increased as time pressure increased. To assess changes in processing quantity with time pressure, we compared the summed total of matching and mismatching ongoing task rates in lower time pressure blocks with the summed total of matching and mismatching rates in higher time pressure blocks. The total processing quantity (matching + mismatching rates) increased from lower to higher time pressure blocks under both low trial-load (Conflicts: $Z = 19.94$, $p < .001$; Non-conflicts: $Z = 16.24$, $p < .001$), and high trial-load conditions (Conflicts: $Z = 31.10$, $p < .001$; Non-conflicts: Z

= 24.76, $p < .001$). Total accumulation decreased, however, from low to high trial load when comparing cells with equal time pressure (Conflicts: $Z = -7.83$, $p < .001$; Non-conflicts: $Z = -6.20$, $p < .001$).

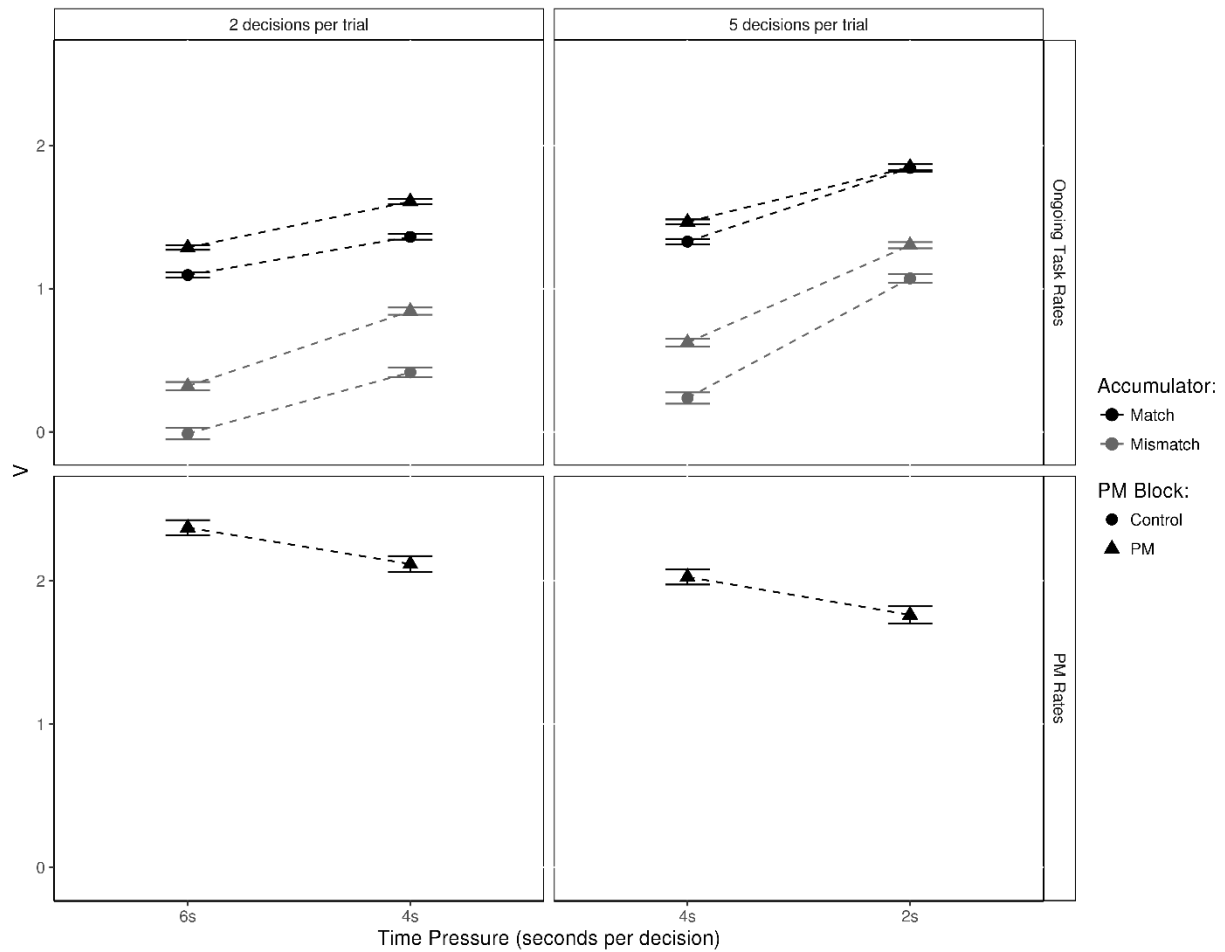


Figure 7. Average ongoing task and PM accumulation rates by time pressure and PM block. Central symbols represent posterior means. Error bars represent ± 1 posterior standard deviation.

We also examined the effects of PM demands on ongoing task processing quantity. The total amount of available ongoing task resources (matching + mismatching rates) increased from control to PM blocks (Conflicts: $Z = 19.58$, $p < .001$; Non-conflicts: $Z = 13.99$, $p < .001$), indicating that more resources were deployed under PM demand. Overall, this indicates that PM

demands and greater time pressure led to greater resource availability; participants deployed more resources under PM load, and as the time available to complete the task decreased.

In contrast to the quantity of ongoing task accumulation rates, rates for the PM accumulator decreased with time pressure under both low trial-load ($Z = -3.78, p < .001$) and high trial-load ($Z = -3.62, p < .001$) conditions, as can be seen in Figure 7 (bottom panel). PM accumulation rates tended to be lower during high versus low trial load when comparing cells with equal time pressure, although this trend did not reach significance ($Z = -1.34, p = .09$). This indicates a trade-off in how resources are allocated or shared between the ongoing and PM tasks, which is affected by changes in time pressure. Specifically, as the ongoing task uses more resources to compensate for additional time pressure demands, it usurps resources that would have otherwise been allocated to the PM task.

Quality. To test for changes in processing quality at different levels of time pressure, we compared the difference between matching and mismatching ongoing task rates in lower time pressure blocks with the difference between matching and mismatching rates in higher time pressure blocks. Figure 7 (top panel) shows rates for ongoing task stimuli (i.e., conflicts and non-conflicts that were not PM targets) across different levels of time pressure. With the exception of low trial-load conflicts, the quality of ongoing task processing was poorer in higher time pressure blocks compared with lower time pressure blocks during both low trial-load (Conflicts: $Z = -1.24, p = .11$; Non-conflicts: $Z = -5.99, p < .001$), and high trial-load conditions (Conflicts: $Z = -6.55, p < .001$; Non-conflicts: $Z = -7.85, p < .001$). Increased time pressure, therefore, reduced the quality of ongoing task processing. Processing quality improved, however, from low to high trial load when comparing cells with equal time pressure (Conflicts: $Z = 2.04, p = .021$; Non-conflicts: $Z = 3.03, p = .001$).

We also tested how ongoing task processing quality varies with PM demands. Consistent with a capacity-sharing effect, the difference between matching and mismatching rates was smaller under PM load compared with control (Conflicts: $Z = -7.63, p < .001$; Non-conflicts: $Z = -5.80, p < .001$). That is, the quality of ongoing task processing was poorer in PM blocks compared with control. There was no significant interaction between time pressure and PM block on ongoing task processing quality.

Taken together, these results are indicative both of an overall increase in the quantity of available resources under time pressure and PM demand, and a degradation in processing quality under time pressure and PM demand. Increasing time pressure produced a decrease in processing quality, which individuals compensated for by increasing the overall rate of information processing (i.e., by working faster). PM load produced similar effects consistent with resource demands: poorer quality processing under PM load, accompanied by an increase in processing quantity. Increasing processing quantity under PM load may have also served to compensate for the slowing associated with higher ongoing task thresholds in PM blocks. We found that higher ongoing task demand led to larger ongoing task quantity but less PM accumulation. This trade-off between ongoing task and PM rates across time pressure conditions demonstrates capacity sharing. That is, individuals shared resources between the ongoing and PM tasks as time pressure increased (i.e., when the available resources were insufficient to compensate for the increased time-pressure demands). The effects of PM load and time pressure on the ongoing task are thus both consistent with a capacity-sharing/resource account.

Reactive Control (PM versus Non-PM Trial Accumulation)

Previous PMDC modeling found ongoing task rates to be lower on PM trials than on non-PM trials due to reactive or stimulus-driven inhibitory control by the PM stimulus detector.

Figure 8 shows rates for ongoing task stimuli that did, and did not, contain a PM target.

Consistent with previous results, ongoing task rates were lower for stimuli containing a PM target compared with stimuli that did not contain a PM target (Conflicts: $Z = 13.36, p < .001$; Non-conflicts: $Z = 13.04, p < .001$). This supports the idea that, when the PM detector identifies a PM target, the accumulation process for the competing ongoing task response is suppressed or inhibited. These findings are consistent with accumulators competing with each other based on their inputs; in the presence of a PM item, evidence accumulation processes for conflicts and non-conflicts are inhibited relative to when a PM item is absent.

The strength of the reactive control effect did not interact with time pressure, although the difference in reactive control strength for non-conflict error accumulation rates did reach significance, with stronger inhibition in higher time pressure blocks (see Table S16). Overall, however, reactive control processes were mostly unaffected by changes in time pressure.

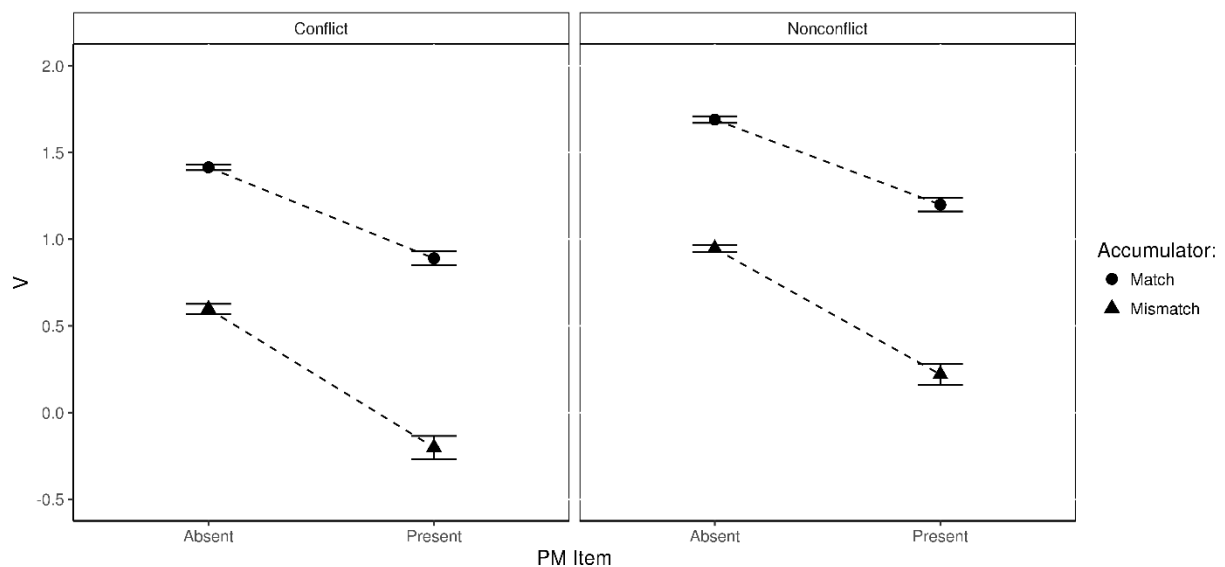


Figure 8. Ongoing task accumulation rates with PM item present versus absent. Central symbols represent posterior means. Error bars represent ± 1 posterior standard deviation.

Model Exploration

In this section we attempt to tease out the individual contribution to accuracy and RT provided by different parameters and mechanisms in the model. It is our aim to give a clearer picture of the relative importance of key mechanisms in accounting for the observed effects. To do this, we first replace the parameter of interest with the average either across control/PM blocks or across time pressure levels (e.g., replacing control block and PM block ongoing task thresholds with the average of the two), and examine the associated misfit of the model with the removed effect relative to the full model with the effect included.

We generated posterior predictions for all PM and time pressure effects for the selected model, and then separately for models with two different sets of parameters averaged over either PM or time pressure blocks: ongoing task rates, and ongoing task thresholds. In the supplementary materials, we include fit graphs comparing the full model to the two versions with proactive (threshold) and reactive (rate) mechanisms systematically turned off.

Here we focus our discussion on time pressure and PM effects and the contribution of each cognitive mechanism. First, we assess how much of the PM accuracy and RT effects were accounted for by the different models, and whether each model adequately fit these effects. Second, we assess how much of the time pressure accuracy and RT effects were accounted for by the different models, and whether each model adequately fit these effects. Finally, we examine the misfit associated with selectively turning off proactive control mechanisms, reactive control mechanisms, or both in terms of fit to overall accuracy and RT.

Time Pressure Effects. The selected model provided a very close account of accuracy differences across different levels of time pressure under both low and high trial load (Figure S6). The selected model also closely fit RT differences between time pressure blocks for both

correct and error ongoing task responses and PM responses. Averaging ongoing task thresholds made the model over-predict the cost to ongoing task accuracy under increased time pressure for conflicts, that is, the model predicted lower conflict accuracy in the high time pressure blocks than was observed. Conversely, the model under-predicted the cost to ongoing task accuracy under increased time pressure for non-conflicts, that is, it predicted higher non-conflict accuracy in the high time pressure blocks than was observed.

Similarly, the averaged-threshold model under-predicted RT differences between higher and lower time pressure blocks for both correct and error ongoing task responses (i.e., it predicted less of a speed-up from lower to higher time pressure than was observed). That is, the primary consequence of lowering ongoing task thresholds under high time pressure was an overall increase in the speed of responding. This provides further evidence that proactive control mechanisms are critical for explaining empirical speed-accuracy trade-offs in ongoing task performance under different levels of time pressure.

Averaging ongoing task rates resulted in a model that under-predicted the cost to ongoing task accuracy under increased time pressure (actually predicting an accuracy advantage for conflicts under higher time pressure instead of the observed cost). Likewise, the averaged-rates model under-predicted RT differences between high and low time pressure blocks for both correct and error ongoing task responses, predicting less of a speed-up from low to high time pressure than was observed. That is, changes in accumulation rates between low and high time pressure conditions increased the speed but reduced the accuracy of ongoing task responses in high time pressure blocks, which is consistent with our findings of increased availability and of reduced quality of ongoing task processing when there is less time available to complete the task(s) at hand.

The averaged-threshold model maintained close fits to PM accuracy, but severely under-predicted PM RT differences between high and low time pressure (Figure S7). In contrast, the averaged-rates model maintained close fits with PM RT, but under-predicted PM accuracy differences between high and low time pressure (i.e., predicted higher PM accuracy under high time pressure than was observed). That is, PM RT was more dependent on ongoing task thresholds than rates, while PM accuracy was more affected by ongoing task rates than thresholds. Taken together, both proactive and reactive control mechanisms appear necessary to account for the full range of accuracy and RT differences in ongoing task and PM performance under different levels of time pressure.

PM Effects. As shown in Figure S8 (supplementary materials), the selected model accounted for the costs to ongoing task accuracy and RT associated with PM demand, providing close fits to the difference between control and PM accuracy for both conflict and non-conflict stimuli, and to correct and error ongoing task RT. Averaging ongoing task thresholds between control and PM blocks made the model over-predict the cost to ongoing task accuracy, more so for conflicts than non-conflicts, and dramatically under-predict the cost to both correct and error ongoing task RT (so much so that the model in fact predicted a PM advantage to ongoing task RT rather than a cost). That is, the model predicts that reallocating resources from the ongoing task to the PM task would have had more of a cost to accuracy if proactive control had not been used. Adopting higher ongoing task thresholds under PM load, therefore, facilitated accurate ongoing task performance under PM load, but was also the primary driver of the PM costs in terms of ongoing task RT. This further indicates that proactive control of thresholds is an important mechanism in explaining PM costs effects, particularly costs to ongoing task RT.

In contrast, averaging ongoing task rates between control and PM blocks caused the model to under-predict costs to ongoing task accuracy (predicting an accuracy advantage for conflicts) and over-predict costs to both correct and error ongoing task RT. That is, changes in accumulation rates between control and PM blocks improved the speed but reduced the accuracy of ongoing task performance under PM load. This is consistent with our findings of increased resource availability and of resources being reallocated from the ongoing task in PM blocks, and it supports the idea that cognitive control of rates is an important mechanism in explaining the differences in ongoing task accuracy and RT that occur under PM demand.

Importance of Cognitive Control for PM Performance. Finally, we examine the overall importance of proactive and reactive control mechanisms in terms of how well the model fits observed accuracy and RT on PM trials. As shown in Figure S9, the full model (which includes both proactive and reactive control mechanisms) provided an almost exact fit to PM accuracy (100.02% of the effect), while slightly over-predicting overall PM RT (RT over-predicted by 1.44%).

In contrast, when proactive control of ongoing task thresholds is removed, and the model only allows for reactive control, we get severe misfit to both overall PM accuracy and RT, with the model under-predicting both effects by a large margin (PM accuracy under-predicted by 17.1% of the effect [i.e., ~60% versus ~72%], PM RT under-predicted by 8.7%). Similarly, when reactive control is removed from the model and only proactive control included, we get less severe but still substantial under-prediction of both PM accuracy and RT (PM accuracy under-predicted by 9.8% of the effect [i.e., ~65% versus ~72%], PM RT under-predicted by 4.4%).

Lastly, when both proactive and reactive control mechanisms are turned off (i.e., the model includes neither mechanism), the model produces the worst overall fit to PM accuracy and

RT, under-predicting both by a larger margin than either the proactive-only or reactive-only models (PM accuracy under-predicted by 30.2% of the effect [i.e., ~50% versus ~72%], PM RT under-predicted by 15.8%). This provides strong evidence that both proactive and reactive cognitive control mechanisms play critical and complimentary roles in accounting for the speed and accuracy of PM task performance.

General Discussion

We used PMDC, a comprehensive computational model of PM, to investigate the cognitive mechanisms underlying performance as a function of PM load and time pressure in a complex dynamic conflict detection task. Our primary aim was to assess the utility of PMDC in modeling a more complex and dynamic task than has been previously studied in the PM literature and gain insight into how cognitive control processes and the quality and quantity of resources drive performance. PMDC provided good fits to ongoing task and PM accuracy and closely accounted for the full distribution of RTs for both ongoing task and PM responses across all of our experimental manipulations. The model accounted for changes in accuracy and RT across different levels of time pressure and fit all PM cost effects. In addition, the model provided out-of-sample predictions of non-response proportions that closely matched empirical non-response data.

Our model fits implicated Braver's (2012) proactive control (increased ongoing task thresholds in PM blocks) and reactive control (inhibited ongoing task accumulation rates to PM items) in PM performance, replicating work with the PMDC model on simple laboratory paradigms in a complex dynamic task environment (Strickland et al., 2018). As discussed in the sections below, we also extend Strickland et al by finding evidence that PM load affected

ongoing task capacity (i.e., the quality of rates was poorer in PM compared to control blocks). Furthermore, manipulating time pressure caused a trade-off between ongoing task accumulation and PM accumulation – higher ongoing task demand led to larger ongoing task quantity but less PM accumulation. That is, while the resources available to the cognitive system increased under time pressure and PM load (increased arousal), with higher time pressure more resources were allocated to the ongoing task at the expense of the PM task, consistent with capacity sharing. Time pressure also reduced the quality of information entering the decision process. Finally, we extended Strickland et al. by showing that PM load-induced proactive control over ongoing task decisions was reduced under higher time pressure.

Capacity Sharing and Resource Availability: Quantity and Quality

Quantity. We found that the total amount of ongoing task information processed (matching + mismatching accumulation rates) was higher in PM blocks than in control blocks, and higher during periods of high time pressure compared with low time pressure. These findings suggest an increase in the amount of resources available to the cognitive system under PM load, and under high time pressure. However, as demands exceeded the capacity limit of the system, resources were reallocated (shared) between the ongoing and PM tasks. This was particularly evident in the trade-off between ongoing task and PM accumulation rates across different levels of time pressure; as the ongoing task used more resources to compensate for additional time pressure demands, it drained resources that would have otherwise been available for PM task processing. This is the critical test of capacity sharing (Navon & Gopher, 1979). Our participants' subjective reports of effort, mental demand, and temporal demand supported this account, and it is consistent with research suggesting that the reallocation of resources amongst competing tasks requires cognitive control (Hockey, 2013; Hockey, Coles, & Gaillard, 1986;

Kleinsorge, 2001; Sanders, 1983; Schmidt, Kleinbeck, & Brockmann, 1984; Spitz, 1988; Wickens, 1986). We also note that modeling this trade-off between concurrent tasks provides another way of revealing capacity-sharing effects that does not rely on comparing performance to a separate ‘control’ block, which is often not feasible for some tasks or experimental settings.

Quality. PMDC indicated that the quality of ongoing task processing was reduced during high time pressure blocks (when there was less time available) compared with low time pressure blocks, in terms of smaller differences in accumulation rate between the matching and mismatching accumulators. This is consistent with modeling studies showing reduced accumulation rates⁷ during high versus low time pressure or when the speed of responding is emphasized (e.g., Heitz & Schall, 2012; Ho, Brown, van Maanen, Forstmann, Wagenmakers, & Serences, 2012; Rae, et al., 2014; Starns, Ratcliff, & McKoon, 2012; Vandekerckhove, Tuerlinckx, & Lee, 2008). These studies have argued that time pressure reduces the quality of information entering the decision process, either because less diagnostic information can be sampled during the shorter response window, or because certain cognitive processes become less efficient at extracting stimulus information. Our findings are also consistent with applied research showing that sufficient time pressure can induce a ‘redline’ or performance bottleneck (e.g., Palada et al., 2018), whereby increased task demands lead to poorer quality processing.

Similarly, quality was poorer in PM blocks compared with control blocks, consistent with participants reallocating attention from ongoing task decisions towards the PM task when under PM demand. This contrasts with previous studies that used simple laboratory tasks and found no

⁷ These studies did not distinguish between accumulation rate quality and quantity, but note that DDM rates correspond most closely to our concept of quality.

capacity effects (Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017, 2018). This may be because the task was more demanding, exhausting the cognitive system such that any more capacity used for PM monitoring required repurposing resources from the ongoing task. It may also be because our PM targets were highly non-focal as they required a different locus of visual attention (callsign) than ongoing task decisions (comparing aircraft location, speed and trajectory) (Einstein & McDaniel, 2005). Overall this suggests that the mixed results regarding capacity sharing and resource allocation in the basic PM literature may be partly due to typical laboratory ongoing tasks being insufficiently demanding to fully engage capacity.

In terms of overt performance, a reduction in processing quality should mainly lead to lower accuracy. At low time pressure, any reduction in accuracy due to the decline in processing quality from control to PM blocks appears to have been washed out by the higher ongoing task thresholds during PM blocks (proactive control), which allowed similar levels of accuracy to be maintained. However, at high time pressure, there was a robust 5% PM block cost to ongoing task accuracy, which was not present at low time pressure. This shows that, at high time pressure and when under PM load, proactive control was not effective in maintaining the accuracy of ongoing task decisions. One important implication of this is that capacity-sharing effects that may not affect performance under low time pressure can produce errors under high time pressure. This also highlights that our modeling approach can detect the action of resource-sharing mechanisms whose overt effects may be masked by compensatory mechanisms (e.g., proactive control), and thus not evident in manifest accuracy and RT.

Taken together, these quantity and quality effects are indicative both of an overall increase in arousal under PM load and time pressure, and a degradation in processing quality

under PM load and time pressure, consistent with capacity-sharing/resource demand effects.

Arousal may increase because the ongoing task becomes subjectively more difficult or engaging when PM demand and time pressure are introduced, leading participants to compensate for greater demands/poorer quality by deploying more resources and changing how resources are allocated between various subtasks. There is a strong neurological basis for this idea (e.g., Banquet, Smith, & Guenther, 1992; Beierholm, Guitart-Masip, Economides, Chowdhury, Düzal, Dolan, Dayan, 2013; Botvinick & Braver, 2015; Chiew & Braver, 2013; Gallistel, 1985); studies have suggested that heightened task demands lead to the potentiation of neural structures related to cognitive processing, for example by increasing the probability that certain neurotransmitters will be released (Jimura, Locke, & Braver, 2010; Niv, Daw, & Dayan, 2007; Schmidt, Lebreton, Clery-Melin, Daunizeau, & Pessiglione, 2012). This has the effect of boosting processing intensity or increasing the signal strength of evidence sampled from stimuli, thus facilitating the flow of information (Kleinsorge, 2001; Wickens, 1986). In the model, such effects would be reflected in changes in the quantity and/or quality of evidence accumulation. An interesting avenue for future research would be to map neural activity to changes in model parameters under different time pressure and task demand manipulations.

Importantly, our findings reveal a potential limitation with using simple laboratory tasks to study PM, which the present work addresses. That is, certain resource availability and allocation effects may not manifest in simple laboratory tasks if the demands of the task rarely exceed the supply of resources (i.e., if the cognitive system is rarely pushed to its capacity limit). In contrast, the current work shows that such effects become readily apparent in more complex paradigms that impose sufficient demands on the cognitive system. Indeed, this issue may have led some of our previous PM research to underemphasize the role of capacity and resource

allocation mechanisms, at least when trying to explain PM in the wild. In contrast to the basic literature, cognitive control and resource management mechanisms play important roles in explaining decision making under PM demand in complex dynamic task environments.

Cognitive Control Processes

Proactive Control. Consistent with proactive control (Braver, 2012), ongoing task thresholds were higher in PM blocks than control blocks, indicating that subjects strategically delayed ongoing task decisions to avoid pre-empting the PM process (Heathcote et al., 2015). The model indicated that this control was the cause of PM cost (i.e., the slower ongoing task RTs in PM blocks). This is consistent with many modeling studies from simple paradigms that indicate costs are driven by increased thresholds to make ongoing task decisions (e.g., Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017, 2018). Further, the model indicated that proactive control over ongoing task thresholds contributed to PM accuracy (about 12% of observed PM responses would have been omitted without proactive control over ongoing task thresholds [i.e., ~60% versus ~72% observed]).

Ongoing task and PM thresholds decreased as time pressure increased, meaning participants set lower thresholds when there was less time available. This is consistent with much choice-RT modeling of the speed-accuracy trade-off which indicates individuals strategically control thresholds in favour of fast responding when time pressure is high (e.g., Dutilh, Wagenmakers, Visser, & van der Maas, 2011; Forstmann et al., 2011; Usher, Olami, & McClelland, 2002). More interesting, however, was the novel finding that PM-induced proactive control interacted with time pressure. Specifically, the difference between control block and PM block ongoing task thresholds got smaller as time pressure increased. This suggests that strategic adjustments to PM demand were smaller under high time pressure compared with low time

pressure. That is, higher levels of time pressure induced a trade-off between raising thresholds to give the PM accumulator more time to finish and lowering thresholds to respond within the trial deadline. This was consistent with the observed drop in PM accuracy in high time pressure blocks, and the observed PM-control differences in ongoing task accuracy and non-response rates, which only reached significance under high time pressure. The drop in PM accuracy under high time pressure indicates that participants were unable to delay responding long enough (due to the short deadline) to prevent ongoing task accumulators out-pacing the PM response. These findings suggest that people may be particularly prone to making PM errors during periods of heightened time pressure, and that the way in which time pressure and PM demands interact could be important to consider in situations in which high PM accuracy is important.

Time pressure also affected threshold bias for the ongoing task (i.e., the amount of evidence required to respond ‘conflict’ versus ‘non-conflict’). Bias shifted from favouring non-conflict responses at lower levels of time pressure to favouring conflict responses at higher levels of time pressure. Under high time pressure, subjects set lower conflict thresholds relative to non-conflict thresholds, which facilitates faster conflict responses but increases the frequency of conflict false alarms, reducing overall accuracy. This finding is in line with Loft et al. (2009), who reported that air traffic controllers apply larger safety margins by shifting bias towards conflict responses when under perceived time pressure. Although this strategy increases the frequency of conflict false alarms, the bias has the effect of ensuring safety over strict accuracy. That non-expert (novice) participants adopted a conflict bias in the current study without an explicit incentive to do so speaks to the face validity of the conflict detection task. That is, participants treat missing an aircraft conflict (i.e., possible passenger fatalities) as worse than making a false alarm (i.e., possibly increasing aircraft travel time) (see Neal & Kwantes, 2009;

Vuckovic et al., 2013, 2014). This finding supports the idea that people can make deliberate strategic adjustments to threshold bias in order to avoid failing to detect critical events (e.g., potential conflicts). Together, these threshold findings highlight the dynamic ways individuals can exert cognitive control over their own decision making in order to balance competing objectives (e.g., safety of aircraft versus efficient traffic flow). An interesting avenue for future research would be to manipulate the relative importance of the ongoing task responses to see if participants adopt different biases based on which types of response are associated with more serious safety consequences.

Reactive Control. Consistent with PMDC's reactive inhibition mechanism, ongoing task accumulation rates were lower on PM trials compared with non-PM trials. That is, rates were lower for stimuli containing a PM target compared with the same stimuli not containing a PM target. This supports the idea that, when a PM item is present, the PM detector inhibits input to the competing ongoing task accumulators. At a strategic level, inhibiting ongoing task rates on PM trials increases the likelihood that the PM accumulator will reach threshold and trigger a PM response before either of the ongoing task accumulators. This finding replicates the reactive control demonstrated by Strickland et al. (2018) in simple laboratory PM tasks, and is in line with theoretical work (Bugg et al. 2013), neurological data (McDaniel et al. 2013), and broader approaches to human error (Norman, 1981; Reason, 1990) that have implicated reactive control in PM. Norman (1981), for example, contends that response inhibition is required for atypical task responses to be able to compete for retrieval with task responses more strongly associated with common environmental cues.

In general, the strength of reactive control was unaffected by changes in time pressure, meaning the level of inhibition did not vary under different levels of time pressure. This lack of

interaction is consistent with Braver's (2012) description of reactive control as an automatic, stimulus-driven control process. Moreover, because reactive control is only active on PM trials, any potential differences in the strength of reactive control can only have a modest effect on overall accuracy in a given block of trials, meaning it would not be an effective mechanism for dealing with changes in time pressure that affect entire blocks of trials.

Trial Load

Trial load (compared across cells with equal time pressure) had several effects. Higher trial load was associated with lower resource availability (quantity) and slightly slower RTs (ongoing and PM). However, higher trial load also led to improved processing quality, slightly lower ongoing task thresholds, and lower subjective reports of effort, mental demand, and temporal demand. This suggests participants found it easier to discriminate conflicts from non-conflicts when they were given 20s to make five decisions, than when they were given 8s to make two decisions, with the lower thresholds and reduced availability possibly reflecting some degree of 'satisficing' via a reduction in the depth of processing (e.g., Donkin, Little, & Houpt, 2014). This pattern of results is consistent with research on temporal discounting (e.g., Ballard, Vancouver, & Neal, 2018), showing that individuals tend to work harder when given more immediate deadlines. Our participants likely perceived having 20s to make five decisions to be a less demanding deadline than having 8s to make two decisions, despite having equal time available per decision on average. One potential explanation for this is the possibility of a fixed start-up time on each trial, which would leave more subsequent processing time on 20-second/5-decision trials compared with 8-second/2-decision trials.

Summary and Further General Implications

In work and everyday life, people must often adapt to PM demands and time constraints in complex and dynamic environments. In the current paper, we present and test a model of the cognitive control and capacity allocation mechanisms that enable PM performance in such settings. Our model provides a cogent explanation for a complex set of empirical phenomena that would be difficult to interpret using standard statistical analyses of observed behaviour, producing several implications for theories of PM. The theoretical insights provided by our model, summarized below, are critical to understanding both PM processes and multiple task performance more generally. For example, the comprehensiveness of PMDC's account is critical for safety-critical work contexts where a characterization of performance fluctuations is necessary to predict and manage risks associated with worst-case performance. The critical 'redline' of workload is the breakpoint of performance as individuals enter the 'overload region', such that they have no capacity to meet further demands and performance deteriorates (Hart & Wickens, 2010). Prior work has typically used self-reported workload to identify the overload region (e.g., Colle & Reid, 2005), but self-reported workload is often dissociated from performance (Yeh & Wickens, 1988). This is because whether performance degrades depends on the strategy that the individual uses (Loft et al., 2007). The PMDC model can be used to explain why performance does or does not decline as workload increases, how performance is protected, and to predict when performance will degrade considerably in a given situation with the use of a specific cognitive strategy or process. By identifying the cognitive redline and the specific cognitive processes responsible for performance decrements under different conditions, our approach can potentially aid the development of work design interventions (e.g., automation,

training, decision aids, interface design) to improve decision making and PM in complex dynamic environments (see Byrne & Pew, 2009).

Our findings demonstrate the importance of using sufficiently demanding tasks to study capacity allocation. As mentioned, our demanding relative judgment task allowed us to detect two robust capacity cost effects: a within-subjects trade-off between PM and ongoing task accumulation rates as task demand (time pressure) increased, and a cost to ongoing task processing quality from PM load, an effect that had been unclear or absent in data from simple laboratory PM paradigms (e.g., Ball & Aschenbrenner, 2017; Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017, 2018). Going forward it will be important for researchers to study PM processes in sufficiently representative tasks to get a clear picture of the processes potentially involved in a given applied task setting (Morrow, 2018; Stokes, 2011).

Overall, our capacity allocation effects are in line with the ecological rationality approach to human behaviour, which contends that decision makers compensate and adapt to environmental constraints in order to achieve the most effective possible outcomes (Brunswick, 1943; Simon, 1956; Todd & Gigerenzer, 2007). Poorer quality ongoing task processing under PM load and time pressure could reflect participants reducing the amount of information processed (i.e., satisficing) by changing their depth of processing (Donkin, Little, & Hout, 2014) or by using heuristics (Gigerenzer & Gaissmaier, 2011). The finding that PM demand and time pressure led to heightened resource availability is in line with observations that arousal levels can change while a person is performing a task (Stearman & Durso, 2016). For example, it is possible that a rapid increase in demand, caused by a sudden increase in the number of stimuli to be processed, can produce an increase in arousal, thus providing additional resources to meet task demands. However, although more resources may become available to meet task demands,

the current modeling results highlight that the cognitive system is limited; when demands exceed the capacity of the system, resources must be reallocated from lower priority tasks to higher priority tasks. The present work is a case in point that researchers cannot rely solely on generalizing findings from simple laboratory paradigms to PM in complex dynamic tasks.

It is important to note, however, that our capacity allocation effects were small in comparison to the proactive and reactive control effects we found here and in previous work (e.g., Heathcote et al., 2015; Strickland et al., 2017; 2018). Moreover, our novel finding that PM-induced proactive control interacted with time pressure (i.e., the size of ongoing task threshold adjustments between control and PM blocks got smaller as time pressure increased, producing costs to PM-block accuracy that were largest under high time pressure) suggests that although attentional capacity is clearly a useful concept for our understanding of PM and ongoing task decision making, it needs to be situated within a cognitive control framework in order to form a complete picture of the processes that drive overt performance.

Limitations and Future Directions

We used an aircraft conflict detection task because it is representative of a broad range of work tasks that require people to remember to perform deferred task actions while making judgements about objects moving on task displays, such as maritime surveillance and submarine track management. The task captured the essential elements of the target phenomenon, because it requires participants to rapidly assess whether moving objects on crossing trajectories will, or will not, violate the applicable separation standard, whilst maintaining an intention to perform a deferred action. Nevertheless, the need for modeling constraints and the need for experimental control dictated that the task be simpler than conflict detection in air traffic control in several ways. For example, in our task, subjects were presented with one stimulus (aircraft pair) at a

time. This is simpler than real air traffic control, in which operators are often required to attend to many aircraft and simultaneously monitor many unfolding potential conflicts with variable onsets and durations. Multi-stimulus environments should be even more demanding than our task and should therefore elicit greater resource reallocation and capacity-sharing effects than seen here. Extending the present modeling to multi-stimulus environments where individuals are free to switch attention among several stimuli represents an interesting avenue for future PMDC work. However, it represents a complex modeling problem, because it removes one of the major constraints on the way participants can adapt to changes in time pressure; they can respond not only by changing threshold, bias and managing cognitive resource availability and allocation, but also how they allocate spatial attention. For example, participants may use a serial strategy, in which they focus on a single stimulus until they make a response, prioritising on the basis of order of onset or deadline. Alternatively, they may use an interleaving strategy, in which many small increments in evidence are made by rapidly switching among stimuli depending on the effort required to move attention, and the dynamically unfolding priority or expected value of the stimuli (see Wickens, Goh, Horrey, Helleberg, & Talleur, 2003).

In our task, subjects made PM responses instead of the typical ongoing task response. This was done because, in many real-world settings, the most serious PM errors are often due to ‘habit capture’, whereby the individual intends to substitute an atypical action for a habitual one, but inadvertently executes the unintended habitual action (Grundgeiger, Sanderson, & Dismukes, 2015; Reason, 1990). However, this paradigm differs from ‘dual-task’ PM (Bisiacchi et al., 2009), in which PM responses are made in addition to the ongoing response. Because proactive and reactive control strategies involve some degree of competition between responses, it is unclear whether similar effects will be present in dual-task paradigms in which subjects are

instructed to make both responses. Heathcote et al. (2015) modeled ongoing task trials in a dual-task PM paradigm and found similar proactive control effects. However, because PM trials have not yet been modeled in dual-task PM, it is unclear whether reactive control processes will operate in a similar way.

A further point, with potential implications for PM theory, is that the current model did not include a mechanism for PM ‘trigger failures’ (Matzke, Love, & Heathcote, 2017) – trials on which the PM intention is not retrieved in the first place (i.e., where the PM accumulator does not enter the race). In PMDC, the PM accumulator always runs on PM trials, meaning PM failures are always due to the PM accumulator losing the race to one of the ongoing task responses. As such, the current modeling framework cannot distinguish between PM failures due to PM evidence not reaching threshold and those due to complete failures to run the PM process. This may be an issue for modeling tasks in which PM errors are expected to be due to retrieval failures, or in designs with features expected to induce forgetting (e.g., spatial context manipulations, see Loft, Finnerty et al., 2011; very low frequency PM targets, see Loft et al., 2016; time-based PM, see Huang, Loft, & Humphreys, 2014; remembering to perform a task after a significant period of interruption, see Wilson, Farrell, Visser, & Loft, 2018). Extending the model to include a trigger-failure mechanism would be relatively straightforward and could allow future work to titrate the two kinds of PM failures.

Finally, future work should look at how cognitive control (proactive/reactive) and resource allocation mechanisms (quantity/quality) are affected by the relative importance of different aspects of the task (Walter & Meier, 2014). One might expect the magnitude of proactive and reactive control effects to be larger when the PM task is prioritized over the ongoing task compared with when the ongoing task or speed of responding are the main priority.

Similarly, one might expect greater ongoing task capacity costs (i.e., reduced ongoing task quality) when the PM task is prioritized (e.g., due to more resources being shunted to the PM task), or heightened arousal (i.e., increased quantity) when the speed of responding is emphasized. As such, a useful follow-up to the present work would be to investigate how individuals adjust cognitive control and resource allocation strategies to meet the demands of different task priorities.

Conclusion

This study applied the PMDC model to account for PM and ongoing task performance in a complex dynamic conflict detection task designed to be representative of many real-world environments in which PM is performed. Drawing on Braver's (2012) dual-mechanisms theory of cognitive control, our analyses showed that resource allocation and cognitive control mechanisms each have an important role in explaining the effects of PM and time pressure on decision making in demanding task environments. Our finding of robust resource availability and reallocation effects associated with PM demand and time pressure, using a cognitively demanding task performed under sufficient time pressure, illustrates the utility of studying PM in representative environments. This work provides the first comprehensive quantitative understanding of how attentional resources and cognitive control mechanisms support PM and ongoing task performance in complex dynamic environments.

Context

This research is related to other works by the authors across several different research fields, including computational modeling of decision making (e.g., Palada, Neal, Tay, & Heathcote, 2018; Trueblood, Brown, & Heathcote, 2014), prospective memory (e.g., Heathcote, Loft, & Remington, 2015; Strickland et al., 2017, 2018), and applied cognition/human factors

(e.g., Loft, 2014; Loft, Chapman, & Smith, 2016; Loft, Finnerty, & Remington, 2011; Loft, Percy, & Remington, 2015). This project bridges gaps between these different fields by using a comprehensive computational modeling framework to test cognitive theory and model latent cognitive control and attentional processes involved in performing a complex task with dynamic time pressure and memory demands.

References

- Anderson, F., Rummel, J. & McDaniel, M. A. (2018). Proceeding with care for successful prospective memory: Do we delay ongoing responding or actively monitor for cues? *Journal of Experimental Psychology: Learning Memory and Cognition*. Retrieved from https://www.researchgate.net/publication/323859395_Proceeding_With_Care_for_Successful_Prospective_Memory_Do_We_Delay_Ongoing_Responding_or_Actively_Monitor_for_Cues
- Australian Transport Safety Bureau. (2010). *Breakdown of separation, VHVXU and A6-EMT, 60KM south east of Mildura. Aviation Occurrence Investigation A0-2009-056. Final*. Canberra, Australia.
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7*. (Vol. 67). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Ball, B. H. & Aschenbrenner, A. J. (2017). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*, 1–9.
- Ball, B. H., & Brewer, G. A. (2018). Proactive control processes in event-based prospective memory: Evidence from intraindividual variability and ex-Gaussian analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(5), 793-811.
- Ballard, T., Vancouver, J. B., & Neal, A. (2018). On the pursuit of multiple goals with different deadlines. *Journal of Applied Psychology*.
- Banquet, J. P., Smith, M., & Guenther, W. (1992). Top-down processes, attention, and motivation in cognitive tasks. In D. Levine & S. Leven (Eds.), *Motivation, Emotion, and Goal Direction in Neural Networks* (pp. 169–208). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor. *Neuropsychopharmacology*, 38(8), 1495.
- Bisiacchi, P. S., Schiff, S., Ciccola, A., & Kliegel, M. (2009). The role of dual-task and task-switch in prospective memory: Behavioural data and neural correlates. *Neuropsychologia*, 47(5), 1362–1373.
- Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, 66, 83–113.
- Bowden, V. K., & Loft, S. (2016). Using memory for prior aircraft events to detect conflicts under conditions of proactive air traffic control and with concurrent task requirements. *Journal of Experimental Psychology: Applied*, 22(2), 211–224.
- Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, 40(1), 70–82.
- Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cerebral Cortex*, 11(9), 825–836.
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113.
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. (2nd ed.). Berkeley: University of California Press.

- Bugg, J. M., McDaniel, M. A., & Einstein, G. O. (2013). Event-based prospective remembering: An integration of prospective memory and cognitive control theories. *The Oxford handbook of cognitive psychology*, 267-282.
- Bugg, J. M., Scullin, M. K., & McDaniel, M. A. (2013). Strengthening encoding via implementation intention formation increases prospective memory commission errors. *Psychonomic Bulletin & Review*, 20(3), 522–527.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523–547.
<https://doi.org/10.1037/0033-295x.97.4.523>
- Byrne, M. D., & Pew, R. W. (2009). A history and primer of human performance modeling. *Reviews of Human Factors and Ergonomics*, 5(1), 225–263.
- Cassey, P., Heathcote, A., & Brown, S. D. (2014). Brain and behavior in decision-making. *PLoS Computational Biology*, 10(7), e1003700.
- Castro, S., Strayer, D. L., Matzke, D., & Heathcote, A. (submitted). Modeling the detection response task. *Conference: Object Perception, Attention and Memory Conference*, Retrieved from
https://www.researchgate.net/publication/324043188_Modeling_the_Detection_Response_Task
- Chiew, K. S., & Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology*, 4, 15.
- Colle, H. A., & Reid, G. B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. *The International Journal of Aviation Psychology*, 15(4), 303–319.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116–127.
- Dambacher, M., & Hübner, R. (2015). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological Research*, 79(1), 83–94.

- Dismukes, K., & Nowinski, J. (2007). Prospective memory, concurrent task management, and pilot error. In A. Kramer, D. Wiegmann, & A. Kirklik (Eds.), *Attention: From Theory to Practice* (pp. 225–236). New York, NY: Oxford University Press.
- Dismukes, K. (2010). Remembrance of things future: prospective memory in laboratory, workplace, and everyday settings. *Reviews of Human Factors and Ergonomics*, 6(1), 79–122.
- Dismukes, K. (2012). Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science*, 21(4), 215–220.
- Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, 16(6), 1129–1135.
- Donkin, C., Brown, S. D., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, 55(2), 140–151.
- Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed-accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1183–1202.
- Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, 35(2), 211–250.
- Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, 17(6), 763–771.
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 717–726.
- Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: multiple retrieval processes. *Current Directions in Psychological Science*, 14(6), 286–290.

- Einstein, G. & McDaniel, M. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and McConnell (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(4), 1082–1088.
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annual Review of Psychology*, 67, 641–666.
- Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: a structural model-based approach. *Journal of Neuroscience*, 31(47), 17242–17249.
- Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab advanced: an air traffic control simulator with realism and control. *Behavior Research Methods*, 41(1), 118–127.
- Fox, J., & Weisberg, S. (2011). Multivariate linear models in R. *An R Companion to Applied Regression*. Los Angeles: Thousand Oaks.
- Gallistel, C. R. (1985). Motivation, intention and emotion: goal-directed behavior from a cognitive-neuro-ethological perspective. In M. Frese & J. Sabini (Eds.), *Goal-directed behaviour: The concept of action in psychology* (pp. 48–65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*, CRC press: Boca Raton, FL.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of Psychology*, 62, 451–482.
- Gobell, J. L., Tseng, C., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research*, 44(12), 1273–1296.
- Grundgeiger, T., Sanderson, P. M., & Dismukes, R. K. (2015). Prospective memory in complex sociotechnical systems. *Zeitschrift für Psychologie*, 222(2), 100–109.

- Grundgeiger, T., Sanderson, P. M., MacDougall, H. G., & Venkatesh, B. (2010). Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied*, 16, 317–334.
- Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: activation/retrieval mode and checking. *International Journal of Psychology*, 38, 245–256.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier.
- Hart, S. G., & Wickens, C. D. (2010). Cognitive workload. *NASA Human Systems Integration Handbook*, 1–17.
- Heathcote, A., Lin, Y-S, Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2018). Dynamic models of choice. *Behavior Research Methods*.
- Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! a delay theory of prospective memory costs. *Psychological Review*, 122(2), 376–410.
- Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, 3, 292.
- Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*, 76(3), 616–628.
- Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, 39(1), 30–47.
<https://doi.org/10.1518/001872097778940597>
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, 53(3), 430–444.

- Ho, T., Brown, S., van Maanen, L., Forstmann, B., Wagenmakers, E.-J., & Serences, J. (2012). The optimality of sensory processing during the speed–accuracy tradeoff. *Journal of Neuroscience*, 32(23), 7992–8003.
- Hockey, R., Coles, M., & Gaillard, A. (1986). Energetical issues in research on human information processing. *Energetics and Human Information Processing*, 3–21.
- Hockey, R. (2013). *The psychology of fatigue: work, effort, and control*. New York, NY: Cambridge University Press.
- Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell us about prospective memory? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 65, 69–75.
- Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 95–117.
- Huang, T., Loft, S., & Humphreys, M.S. (2014). Internalizing versus externalizing control: different ways to perform a time-based prospective memory task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40, 1064–1071.
- Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences*, 107(19), 8871–8876.
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall, Inc: Englewood Cliffs, NJ.
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent–trait approach. *Psychometrika*, 75, 70–98.
- Kleinsorge, T. (2001). The time course of effort mobilization and strategic adjustments of response criteria. *Psychological Research*, 65(3), 216–223.

- Lerche, V., & Voss, A. (2017). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*. <https://doi.org/10.1007/s00426-017-0945-8>
- Liu, C. C., & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision Research*, 61, 107–114.
- Loft, S. (2014). Applying psychological science to examine prospective memory in simulated air traffic control. *Current Directions in Psychological Science*, 23(5), 326–331.
- Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, 15(2), 106–124.
- Loft, S., Chapman, M., & Smith, R. E. (2016). Reducing prospective memory error and costs in simulated air traffic control: external aids, extending practice and removing perceived memory requirements. *Journal of Experimental Psychology: Applied*, 22, 272–284.
- Loft, S., Finnerty, D., & Remington, R. W. (2011). Using spatial context to support prospective memory in simulated air traffic control. *Human Factors*, 53(6), 662–671.
- Loft, S., Humphreys, M., & Neal, A. (2004). The influence of memory for prior instances on performance in a conflict detection task. *Journal of Experimental Psychology: Applied*, 10(3), 173–187.
- Loft, S., Neal, A., & Humphreys, M. (2007). The development of a general associative learning account of skill acquisition in a relative arrival-time judgement task. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 938–959.
- Loft, S., Percy, B., & Remington, R. W. (2015). Varying the complexity of the prospective memory decision process in an air traffic control simulation. *Zeitschrift Für Psychologie*, 219, 77–84.

- Loft, S., & Remington, R. W. (2010). Prospective memory and task interference in a continuous monitoring dynamic display task. *Journal of Experimental Psychology: Applied*, 16(2), 145–157.
- Loft, S., & Remington, R. W. (2013). Wait a second: brief delays in responding reduce focality effects in event-based prospective memory. *Quarterly Journal of Experimental Psychology*, 66(7), 1432–1447.
- Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en-route air traffic control: critical review and broader implications. *Human Factors*, 49(3), 376–399.
- Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: external aids that signal when to act. *Journal of Experimental Psychology: Applied*, 17(1), 60–70.
- Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, 19(3), 254–265.
- Loft, S., & Yeo, G. (2007). An investigation into the resource requirements of event-based prospective memory. *Memory & Cognition*, 35(2), 263–274.
- Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: general and special theories of an act of control. *Psychological Review*, 121(1), 66–95.
- Loukopoulos, L. D., Dismukes, K., & Barshi, I. (2009). *The multitasking myth: handling complexity in real-world operations*. Ashgate Publishing, Ltd: Surrey, England.
- Marsh, R. L., Hicks, J. L., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 652–659.

- Matzke, D., Love, J., Heathcote, A. (2017). A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behavior Research Methods*, 49(1), 267–281.
- McDaniel, M. A., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable neural routes to successful prospective memory. *Psychological Science*, 24(9), 1791–1800.
- Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Reason*, 4(2), 61–64.
- Morrow, D. G. (2018). Publishing papers that matter. *Journal of Experimental Psychology: Applied*, 24(1), 1–2.
- Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 37(7), 2335–2343.
- Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, 91(2), 216.
- Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, 86(3), 214–255.
- Neal, A., Ballard, T., & Vancouver, J. B. (2017). Dynamic self-regulation and multiple-goal pursuit. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 401–423.
- Neal, A., & Kwantes, P. J. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors*, 51(2), 164–180.

- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3), 507–520.
- Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review*, 83(6), 508–510.
- Norman, D., A. (1981). Categorization of action slips. *Psychological Review*, 88, 1–15.
- Osth, A. F., Bora, B., Dennis, S., & Heathcote, A. (2017). Diffusion vs. linear ballistic accumulation: different models, different conclusions about the slope of the zROC in recognition memory. *Journal of Memory and Language*, 96, 36–61.
- Palada, H., Neal, A., Tay, R., & Heathcote, A. (2018). Understanding the causes of adapting, and failing to adapt, to time pressure in a complex multi-stimulus environment. *Journal of Experimental Psychology: Applied*.
- Palada, H., Neal, A., Vuckovic, A., Martin, R., Samuels, K., & Heathcote, A. (2016). Evidence accumulation in a complex task: making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied*, 22(1), 1–23.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140–160.
- Posner, M., & Petersen, S. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13(1), 25–42.
- Raab, O. H. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences*, 24, 574–590.
- Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1226–1243.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922.

- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333–367.
- Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: accounting for both magnitude and difference effects. *Cognitive Psychology*, 103, 1–22.
- Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.
- Rummel, J., Smeekens, B. A., & Kane, M. J. (2017). Dealing with prospective memory demands while performing an ongoing task: shared processing, increased on-task focus, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1047–1062.
- Sanders, A. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53(1), 61–97.
- Schmidt, K-H., Kleinbeck, U., & Brockmann, W. (1984). Motivational control of motor performance by goal setting in a dual-task situation. *Psychological Research*, 46(1-2), 129–141.
- Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The dynamic multiprocess framework: evidence from prospective memory with contextual variability. *Cognitive Psychology*, 67, 55–71.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 347–361.
- Smith, R. (2010). What costs do reveal and moving beyond the cost debate: reply to Einstein and McDaniel (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1089–1095.

- Soar, J., Nolan, J. P., Böttiger, B. W., et al. (2015). European Resuscitation Council Guidelines for Resuscitation 2015. *Resuscitation*, 95, 100–147.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.
- Spitz, G. (1988). Flexibility in resource allocation and the performance of time-sharing tasks. *Proceedings of the Human Factors Society Annual Meeting*, 32, 1466–1470.
- Starns, J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, 64(1–2), 1–34.
- Stearman, E. J., & Durso, F. T. (2016). Vigilance in a dynamic environment. *Journal of Experimental Psychology: Applied*, 22(1), 107–123.
- Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision making, cue recognition, and expertise under time pressure. *Naturalistic Decision Making*, 183–196.
- Stokes, D. E. (2011). *Pasteur's quadrant: basic science and technological innovation*. Washington, DC: Brookings Institution Press.
- Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(10), 1616–1629.
- Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: a theory of decision control in event-based prospective memory. *Psychological Review*, 21, 3.
- Swets, J. A. (1973). The relative operating characteristic in psychology: a technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, 182(4116), 990–1000.

- Swets, J. A., Green, D. M., Getty, D. J., & Swets, J. B. (1978). Signal detection and identification at successive stages of observation. *Perception & Psychophysics*, 23(4), 275–289.
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: ecological rationality. *Current Directions in Psychological Science*, 16, 167–171.
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, 121(2), 179–205.
- Turner, B., Sederberg, P., Brown, S., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, 18(3), 368–384.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. (2008). A Bayesian approach to diffusion process models of decision-making. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 1429–1434.
- Verbruggen, F. & Logan, G. D. (2009). Proactive adjustments of response strategies in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 835–854.
- Vuckovic, A., Kwantes, P. J., Humphreys, M., & Neal, A. (2014). A sequential sampling account of response bias and speed–accuracy tradeoffs in a conflict detection task. *Journal of Experimental Psychology: Applied*, 20(1), 55–68.
- Vuckovic, A., Kwantes, P. J., & Neal, A. (2013). Adaptive decision making in a dynamic environment: a test of a sequential sampling model of relative judgment. *Journal of Experimental Psychology: Applied*, 19(3), 266–284.
- Walter, S., & Meier, B. (2014). How important is importance for prospective memory? a review. *Frontiers in Psychology*, 5, 657.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67–85.

- Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and Performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.
- Wickens, C. D. (1986). The effects of control dynamics on performance. In K. R Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. 2. Cognitive Processes and Performance* (pp. 1–60). Oxford, England: John Wiley & Sons.
- Wickens, C.D., Goh, J., Horrey, W.J., Helleberg, J.H., & Talleur, D.A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors*, 45(3), 360–380.
- Wilson, M. S., Farrell, S., Visser, T. A. W., & Loft, S. (2018). Remembering to execute deferred tasks in simulated air traffic control: The impact of interruptions. *Journal of Experimental Psychology: Applied*.
- Yeh, Y.-Y. & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, 30(1), 111–120.